

PO0001: Gene Editing/CRISPR

A Modular Cloning Toolkit for Genome Editing in Plants

Florian Hahn¹, Andrey Korolev², Laura Sanjurjo Loures¹ and **Vladimir Nekrasov**¹, (1)Rothamsted Research, Harpenden, United Kingdom, (2)John Innes Centre, Norwich, United Kingdom

The modular cloning (MoClo), based on the Golden Gate (GG) method, has enabled development of cloning systems with standardised genetic parts, e.g. promoters, coding sequences or terminators, that can be easily interchanged and assembled into expression units, which in their own turn can be further assembled into higher order multigene constructs. Here we present an expanded cloning toolkit that contains modules encoding a variety of CRISPR/Cas-based nucleases and their corresponding guide RNA backbones. Among other components, the toolkit includes a number of promoters that allow expression of CRISPR/Cas nucleases (or any other coding sequences) and their guide RNAs in monocots and dicots. As part of the toolkit, we present a set of modules that enable quick and facile assembly of tRNA-sgRNA polycistronic units without a PCR step involved. We believe the toolkit will contribute towards wider adoption of the CRISPR/Cas genome editing technology and modular cloning by researchers across the plant science community.

PO0003: Gene Editing/CRISPR

Promoter Region CRISPR/Cas9 Genome Editing to Improve GABA Accumulation in Tomato Fruits

Pietro Gramazio¹, Johan Hunziker¹, Yoshihito Shinozaki¹ and Hiroshi Ezura², (1)University of Tsukuba, Tsukuba, Japan, (2)Faculty of Life and Environmental Sciences, University of Tsukuba, Tsukuba, Japan

Hypertension is a primary risk factor for cardiovascular disease, the latter causing 7.5 million deaths every year. The γ -aminobutyric acid (GABA) is a non-proteinogenic amino acid that has shown effective in lowering the blood pressure of hypertensive patients. Intake of GABA through the daily diet may help reduce their symptoms and might be an effective way to prevent hypertension.

Tomato is one of the most produced vegetables worldwide and daily consumed in a plethora of recipes. A part of many beneficial compounds, tomato contains also high levels of GABA. However, GABA concentration is higher in mature tomato green fruits and rapidly decreases in ripe fruits due to two main mechanisms in the GABA pathway. In fact, GABA is synthesized from glutamate by glutamate decarboxylase (*GAD*) and reversibly converted to succinic semialdehyde by GABA transaminase (*GABA-T*).

A high-GABA content Micro-Tom line was developed deleting the autoinhibitory domain of *SIGAD* gene by introducing a stop codon using a conventional CRISPR/Cas9 approach (Nonaka et al., 2017). However, that approach would not be suitable to edit also *SIGABA-T* gene since it would lead to severe dwarfism and affected vegetative and reproductive growth (Koike et al., 2013).

Our study aims at further increasing the GABA content in tomato applying an alternative CRISPR/Cas9 approach by targeting the promoter region of *SIGABA-T*. In that way, using multiple target gRNAs, we expect to produce novel cis-regulatory alleles that may provide a continuum of variation in *SIGABA-T* gene expression and consequently a better balance between GABA accumulation and plant development.

PE0004: Gene Editing/CRISPR

CRISPR Cas9-Mediated Depletion of Ribosomal RNA Sequences from Different Bacteria

Azeem Siddique¹, Gaia Suckow², Hallie Sussman², Dhanya Ramachandran², Nils Homer³, Steve Head², Phil Ordhoukhanian² and Keith Brown¹, (1)iGenomX, Carlsbad, CA, (2)The Scripps Research Institute, La Jolla, CA, (3)Fulcrum Genomics, Somerville, MA

Although messenger RNA (mRNA) is the focus of much RNA research, it constitutes a relatively small fraction of total RNA in a cell. Most cellular RNA is ribosomal RNA (rRNA) and removal of this RNA is desirable in many

RNAseq studies in order to maximize capacity of the sequencing instrument and reduce costs. There are a number of methods to achieve this: bead-bound poly-T oligonucleotides can be used to enrich for poly-A tailed mRNA, or rRNA can be removed by hybridization to short complementary sequences and capture by beads or hybridization to such sequences and subsequent RNase H treatment. However, rRNA removal methods tend to be expensive and often require large amounts of RNA as input.

We have employed the CRISPR Cas9 double-stranded DNA endonuclease to develop a different method of rRNA removal for RNAseq studies. It is a post-NGS library rRNA removal that can be used on multiplexed samples. The CRISPR Cas9 enzyme targets DNA for cleavage in a site-specific manner when it is complexed with CRISPR RNA (crRNA). The latter directs Cas9 to the DNA sequence with complementarity to a 20 nucleotide target-specific sequence in the crRNA and Cas9 cleaves at that site. In this CRISPR-mediated rRNA depletion method, a library of crRNAs are designed based on targets in the 16S and 23S rRNA genes. Completed RNAseq libraries produced from total cellular RNA are treated with Cas9 complexed to the library of crRNA guides. NGS library molecules containing rRNA sequences are cleaved and, thus, are no longer substrates for PCR or sequencing. After Cas9 treatment, the library is PCR amplified, size selected and sequenced. Here we show data from the CRISPR treatment of rRNA from RNAseq libraries prepared from three different bacteria: *Escherichia coli*, *Methylococcus capsulatus* and *Staphylococcus aureus*.

PO0005: Gene Editing/CRISPR

Using CRISPR to Create Covercress – a Novel Winter Crop with Canola-like Composition That Helps to Prevent Soil Erosion

Tim Ulmasov, CoverCress Inc., St. Louis, MO

To mitigate negative effects of modern agriculture on the environment and to contribute to renewable sources of food and energy, there is a strong need for the development of new crops that are suitable for both human and animal consumption and biofuel production, but which do not compete for land area with food crops. Pennycress (*Thlaspi arvense*), a member of the Brassicaceae, is a close relative of Arabidopsis and canola and very amenable for improvement due to its small, diploid genome (<4X Arabidopsis), ease of transformation (floral dip), short cycle (90-110 days) and presence of highly efficient gene editing system (demonstrated ability to convert four loci into new homozygous states in one generation, free of the marker or CRISPR machinery). It can be used to produce vegetable oil with attractive fatty acids profile that can also be used as a feedstock for biodiesel or jet fuel, as well as protein-rich seed meal that serves as animal feed. Pennycress seeds have high oil content (30-35%) with unique characteristics, such as superior cold-flow properties resulting from the lowest saturated fat content among commercially available plant-based oils (<4%).

The winter annual life cycle of pennycress enables planting it after the fall maize harvest in the Midwest and harvesting prior to soybean planting in late May. Therefore, pennycress cultivation will require neither displacement of existing food crops, nor introduction of new land into agricultural use. Moreover, grown as a winter annual, pennycress serves as a cover crop, protecting soil from erosion, capturing excess nitrogen remaining after corn fertilization, and reducing leakage of nutrients into waterways during late fall and winter.

Using conventional breeding and advanced genome editing technologies, we were able to rapidly convert wild pennycress into CoverCress, a canola-like crop with unique properties, such as a combination of valuable oil and meal composition with a potential to save the environment and generate significant new value for the farmer and entire value chain. Planned for a 2021 launch, the entry market consists of Midwestern corn/soybean farmers, representing about 35M acres of prime farmland that today lie bare following the corn harvest and before soybean planting.

PE0006: Gene Editing/CRISPR

Mutations in the Promoter and CDS of FAD2 Induced By CRISPR-Based Gene Editing in Peanut

Anjanasree K. Neelakandan¹, David A. Wright¹, Sy M. Traore², Xiangyu Chen³, Fengzhen Liu⁴, Martin H. Spalding¹ and **Guohao He**², (1)Iowa State University, Ames, IA, (2)Tuskegee University, Tuskegee,

AL, (3)Fujian Academy of Agricultural Sciences, Fuzhou, China, (4)State Key Laboratory of Crop Biology, Shandong Key Laboratory of Crop Biology, Shandong Agricultural University, Tai'an, China

Fatty acid desaturase 2 (FAD2) enzyme catalyzes the conversion of oleic acid to linoleic acid in peanut seeds. High linoleic acid content contributes to rancid flavors and odor in stored products. Natural mutations in the coding region of FAD2 result in the decrease of its enzymatic activity, leading to an increase in the content of oleic acid, which is beneficial to consumers and industries. The aim of this study was to induce mutations in the FAD2 coding region targeted by a gRNA using CRISPR/Cas9 technology. In parallel, we have designed gRNAs targeting the distal (RY repeat motif) and proximal (2S seed protein motif) regions of the FAD2 promoter. Because RY element and 2S seed protein motifs are implicated in the regulation of seed specific gene expression and linoleic acid may be required for healthy plant growth, gene editing the FAD2 promoter may generate seed with reduced linoleic acid while maintaining normal fatty acid profiles in other plant tissues. Our results showed that indel mutations were generated in the CDS and promoter regions of both homeologous *ahFAD2A* and *ahFAD2B* alleles by CRISPR-based gene editing. In the promoter region, the mutagenesis rate was higher at the distal region than at the proximal region. The induced mutations could reduce the expression of FAD2 genes though phenotypes need to be verified.

PO0007: Gene Editing/CRISPR

Reducing Food Waste By Improving Shelf Life of Perishable Produce Using Patented Gene Technology

Venkatramana Pegadaraju, Agribody Technology, San Diego, CA

Venkatramana Pegadaraju & Jerry Feitelson

Technologies, 12526 High Bluff Dr. #300, San Diego, CA, 92130-2067

Abstracts

Globally 1/3 global food produced is wasted each year, which accounts for 1.3B tons and a loss of \$1 Trillion in retail value. Nearly, 40-50% of the wasted food in roots and tuber, fruits and vegetable, with ~70% of losses in developed countries and ~30% in the emerging countries. Wasted food and its effects on people, the environment, and the economy has become a major topic of international conversation, and for good reasons. Consumers demand access to fresh food that can nourish them in an environmentally friendly manner. Food wastage can be significantly reduced by improving the shelf-life of perishable produce. A combination of superior plant genetics along with existing conventional chemical and physical solutions will offer an attractive solution to this complex global challenge.

Technology has a unique strategy that offers a compelling value proposition to our customers. As opposed to conventional approaches, our methods involve knocking down of a critical regulatory gene controlling plant cell fate that causes perishable crops to quickly decay. Since, this a built-in genetic mechanism, it offers multi-fold value in maintaining freshness of the produce from farm to fork. The efficacy of this technology is consistently proven in multiple crops under both greenhouse and field conditions. Currently, several licensing and co-development projects are underway in potato, tomato, sweetcorn, lettuce, canola, rice and others. Results from these various studies will be discussed.

PO0009: Gene Editing/CRISPR

Developing Novel Potato Properties Using Genome Editing

Diem NP Ly^{1,2}, Stephen P Milroy^{2,3}, Sadia Iqbal² and Michael GK Jones^{2,3}, (1)College of Science, Health, Engineering and Education, Murdoch University, Perth, WA, Australia, (2)WA State Agricultural Biotechnology Centre, Murdoch University, Perth, WA, Australia, (3)Potato Research WA, Murdoch University, Perth, WA, Australia

Potato (*Solanum tuberosum* L.) is the fourth most important food crop worldwide and a significant component of food security. Conventional potato breeding is time-consuming and challenging because as a heterozygous polyploid, it can take up to 20 years to generate a new commercial variety.

A major use of potatoes is their processing into crisps and French Fries. Potatoes which accumulate reducing sugars and exhibit browning after frying are rejected, causing economic losses and waste. The formation of acrylamide when potatoes are deep-fried also presents a consumer health risk, because acrylamide is a potential neurotoxin/carcinogen. One solution to reduce browning and acrylamide formation is to down-regulate two genes linked to these characteristics: *vacuolar invertase (VInv)* and *asparagine synthetase 1 (AS1)*. This project aims to upgrade existing potato cultivars using genome editing (CRISPR/Cas9) to down-regulate expression of the *VInv* and *AS1* genes, to reduce cold-induced sweetening and acrylamide formation in fried potatoes, without otherwise changing the variety.

Target genes have been cloned and sequenced, and single guide (sg) RNAs have been made to cut two regions of each gene. The Cas9 and sgRNAs have been delivered into plant cells either as expression vectors using *Agrobacterium tumefaciens* or directly by particle bombardment as a ribonucleoprotein (RNP) complex. The latter aims to generate DNA-free genome-edited plants that will not be regulated as GMOs. In this on-going project, a range of shoots have been regenerated following both treatments and are undergoing analysis for changes in target gene sequences and improved properties for human consumption.

PE0010: Gene Editing/CRISPR

Genome Editing in Haploid Wheat Tissue for Trait Improvement

Colin Hiebert and **Andriy Bilichak**, Morden Research and Development Centre, Agriculture and Agri-Food Canada, Morden, MB, Canada

Recent advances in genome engineering technologies based on designed endonucleases (DE) allow specific and predictable alterations in plant genomes to generate value added traits in crops of choice. Current methods for delivering genome editing reagents (Cas9/gRNA) into plant cells are based on traditional genetic transformation methods that result in stable integration of the nuclease in the genome. Such transformation events can result in unintentional collateral effects, such as disruption of unrelated loci, off target effects and potential regulatory burden of resulted cultivars even following transgene segregation. Bread wheat is allohexaploid (AABBDD, $2n=6x=42$), therefore, making it challenging to generate null mutants in cost efficient and timely matter. Recently, we have published an approach for genome editing in haploid microspores and embryos by direct delivery of ZFN proteins using cell-penetrating peptide complexes (Bilichak et al., 2019). Here, we present an extension to this technique by optimization of conditions for efficient genome editing of haploid embryos obtained from wide cross hybridization between corn and wheat. Both approaches were explored – conventional delivery of plasmid coding for CRISPR/Cas9 construct, as well as delivery of Cas9/gRNA ribonucleic proteins (RNPs). Developed method can potentially streamline the generation of null wheat mutants in non-transgenic and cost-efficient way.

PO0011: Gene Editing/CRISPR

Development of an *In Planta* Genome Editing System in Wheat

Haruyasu Hamada¹, Yuya Kumagai², Yuelin Liu², Yozo Nagira¹, Ryuji Miki¹, Naoaki Taoka¹ and **Ryozo Imai²**, (1)KANEKA Corporation, Takasago, Japan, (2)NARO Institute of Agrobiological Sciences, Tsukuba, Japan

Recently we have developed an *in planta* transformation system (iPB) that relies on particle bombardment-based gene transfer into shoot apical meristems (SAMs) in wheat (*Triticum aestivum* L.) (Hamada et al. Sci Rep 2017). SAMs contain a subepidermal cell layer, L2, from which germ cells later develop during floral organogenesis. Since the iPB method does not require callus culture or regeneration steps, it is applicable to many elite varieties that are recalcitrant to tissue culture-based conventional transformation methods. Here, we report application of the iPB method to *in planta* genome editing. Gold particles coated with plasmids expressing CRISPR/Cas9 components were bombarded into SAM-exposed embryos of imbibed mature seeds. The embryos were grown to adult plants and tested for mutations using cleaved amplified polymorphic sequence (CAPS) analysis. 5.2% of the bombarded plants

carried mutant alleles, and 1.4% of the bombarded plants contained mutations inherited to the next generation (T1). Genotype analysis of T1 plants identified plants homozygous for the three homoeologous genes. These plants showed no detectable integration of the Cas9 and guide RNA genes, indicating that transient expression of CRISPR/Cas9 created the mutations. As for the next step, we introduced CRISPR/Cas9 ribonucleoprotein using the same procedure. Mutations were detected in 6-8% and 1-4% of bombarded plants in T0 and T1 generations, respectively. Together, our data demonstrated that the iPB method can be used to achieve DNA-free *in planta* genome editing in wheat and suggested possible applications to other recalcitrant plant species and variations.

PE0012: Gene Editing/CRISPR

Precise Genome Editing in Barley Using Ribonucleoprotein-Complexes

Martin Becker, Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Seeland/OT Gatersleben, Germany and Goetz Hensel, Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Seeland / OT Gatersleben, Germany

Targeted mutagenesis in plants using sequence specific endonucleases (SSE) like CRISPR/Cas technology has been shown in a wide range of species. Frontiers of this technique are still the lack of predictability of the outcome, since SSE-introduced double-strand breaks are resulting in insertions and deletions (InDels), which are itself not predictable. Using the homology-directed repair (HDR) mechanism in the cell a predicted allele-exchange can be introduced into the loci by providing a synthetic repair template including the desired gene modification. One way to achieve this precise allele exchange is the use of ribonucleoprotein complexes (RNP). A synthetic sequence-specific gRNA and a Cas protein are assembled *in vitro* and transferred into the cell without integration of foreign DNA sequences.

Here we present data comparing SpCas9 and AsCas12a endonuclease in barley epidermal leaf cells because of their different features. In order to interfere with the ratio between preferred non-homologous end joining (NHEJ) and homology-directed repair (HDR), either RNAi or chemical inhibitors will be used to repress key genes involved in NHEJ. Until now RNAi constructs targeting the well-known NHEJ-related genes *Ku70*, *Ku80* and *Ligase IV* were generated and were used for transient and stable integration in barley. To facilitate easy detection of homology-directed genomic modifications we make use of transgenic barley *gfp* mutants. After DSB-induction and application of the custom repair-template, Gfp fluorescence can be used as readout.

PO0013: Gene Editing/CRISPR

Genome Editing of BS1 Locus in Chickpea (*Cicer arietinum* L.) for Increased Seed Size

Shallu Thakur, Alok Das, Meenal Rathore and Narendra P Singh, ICAR-Indian Institute of Pulses Research, Kanpur, India

Chickpeas (*Cicer arietinum* L.) are important grain legumes vital to nutritional security and foster sustainable agriculture. The most important agronomic trait in chickpea breeding is yield, which includes increased seed size and weight. The locus BIG SEEDS 1 (BS1) is a negative regulator in controlling size of plant organs, including seeds, pods and leaves in the legume *Medicago truncatula*. The deletion and down-regulation of this gene leads to increase in plant organs size including leaf and seed. Comparative studies of BS1 sequences of *desi* chickpea (ICCV 4958) and *kabuli* chickpea (CDC Frontier) at protein level indicated difference in protein size with conserved TIFY domain. To document the natural variants of the BS1 locus, present in chickpea germplasm a subset of 200 genotypes (*desi*, *kabuli* and *wild*) varying in seed size and weight were selected and 9 primer sets covering the entire predicted gene were designed using primer walking concept (Primer 3 software). Association of genotypic differences to phenotype were documented and unique genomic variants/sites were identified for genome editing. A total of 13 single-guide RNAs (sgRNAs) specific to BS1 locus was designed using software tools *viz.* CHOP CHOP software (<http://chopchop.cbu.uib.no/>) and CCTOP (<https://crispr.cos.uni-heidelberg.de/>) using default parameters based on genomic location and their possible off targets effects. *In vitro* screening of 13 sgRNAs, indicated only 8 could be efficiently cleaved by the Cas9. Designed sgRNAs were sub-cloned in the binary vector and employed for *Agrobacterium tumefaciens* mediated genetic transformation of chickpea.

PE0014: Gene Editing/CRISPR

Golden Lettuce: Increasing Carotenoid Content in Lettuce

Amber Marie Flores, University of California, Davis, Davis, CA

Lettuce (*Lactuca sativa*) is one of the most widely consumed vegetables in the world and is one of the leading market-value crops in the US (Parr, Bond, and Minor 2019). However, it is often characterized as being relatively nutritionally inert, lending itself to much needed improvement of its nutritional profile. Carotenoids are naturally occurring pigments in plants that confer significant health benefits, including anti-cancer and prevention of eye and cardiovascular disease to those who consume them (Saini, Nile, and Park 2015). Lettuce, on average, contains less than half of the common carotenoids that are often found in other commonly consumed vegetables (Saini, Nile, and Park 2015). With the implementation of modern gene editing techniques, the ability to design lettuce to contain desirable traits such as increased carotenoids has become increasingly feasible. Previous studies in *Arabidopsis thaliana* (Arabidopsis) and tobacco (*Nicotiana tabbacum*), have demonstrated increased carotenoid content by modulating the expression of core regulators of the endogenous carotenoid pathway. Until recently, modulation of metabolic pathways was only possible through the use of transgenics or time-consuming QTL mapping and traditional breeding. Now, gene expression can be rapidly modulated using CRISPR/Cas9 to induce novel cis-regulatory elements by targeting the promoter regions of key regulatory genes (Rodríguez-Leal et al. 2017). This study proposes a novel approach to increasing carotenoid content of lettuce without the use of transgenics or prolonged traditional breeding efforts by targeting CRISPR/Cas9 to the promoter regions of key regulatory genes in the carotenoid biosynthesis pathway to create new positive cis-regulatory elements.

PO0015: Gene Editing/CRISPR

Applications of Precision Breeding and Genome Editing for Trait Improvement in Vitis.

Sadanand Dhekney¹, Behnam Khatabi¹, Xinhua Dai² and Yunde Zhao³, (1)University of Maryland Eastern Shore, Princess Anne, MD, (2)University of California San Diego, San Diego, CA, (3)Section of Cell and Developmental Biology, University of California, San Diego, San Diego, CA

Conventional breeding of grapevine for trait improvement such as disease resistance and quality characteristics has limited applications due to the heterozygous nature of the grape genome, long juvenile period, and complex genetic control of enological traits. Precision breeding is defined as an approach to plant genetic improvement that transfers only specific traits among sexually compatible relatives via the relatively stable mitotic cell division pathway in order to avoid the significant genetic disruption imposed upon conventional breeding by meiosis. Significant progress has been made in the development of grapevine cell culture, regeneration and gene insertion systems. This coupled with *Vitis* genome sequencing has increased our knowledge about the use of grape-derived genes and genetic elements for improvement of specific traits without disrupting existing desirable characteristics. Advances made to enable precision breeding are currently being extended to optimize grapevine genome editing using CRISPR/Cas9 systems. In the current study, a grapevine phytoene desaturase 1 (PDS1) gene was targeted using CRISPR/Cas9. Guide RNAs (gRNA) targeting the PDS1 gene were placed in a binary construct under the control of a strong *Arabidopsis* promoter along with a hygromycin resistance gene. Embryogenic cultures were co-cultivated with *Agrobacterium* harboring the PDS1 constructs. Transformed embryogenic cultures were grown on selection media to recover edited embryo lines and plants. Phenotypic analysis of the edited plant lines was recorded to evaluate editing efficiency. Targeting of the PDS1 gene resulted in embryo and plant phenotype with a bleached appearance. While embryo and plant lines were albino and completely devoid of chlorophyll, some lines exhibited mosaic leaf and stem patterns. Grapevine lines edited for the PDS1 exhibited a slow growth compared to non-edited control lines and could not survive following transfer to soil and acclimation under conditions of high humidity. We are currently evaluating the possibility of targeting traits for disease resistance and quality improvement using CRISPR/Cas9. The development of such systems can overcome limitations encountered in conventional breeding and enable rapid improvement of existing grapevine cultivars for abiotic/biotic stress tolerance and qualitative traits.

PE0016: Gene Editing/CRISPR

Generation of Lactoferrin Gene Knock-in Pigs Improves the Antibacterial Activities of Sow Milk Via CRISPR/Cas9 Mediated Homologous Recombination

Xiaosong Han, Yang Gao, Youcai Xiong, Guanglei Li, Changzhi Zhao, Shengsong Xie, Xinyun Li, Jinxue Ruan and Shuhong Zhao, Key Laboratory of Agricultural Animal Genetics, Breeding, and

Reproduction of the Ministry of Education & Key Laboratory of Swine Genetics and Breeding of Ministry of Agriculture, Huazhong Agricultural University, Wuhan, China

Pre-weaning piglet mortality is a major economic and welfare issue in pig industry. Sow milk, especially colostrum, contains abundant immune active materials, is the only energy sources of newborn piglets. The bioactive components can provide passive immunity to new born pigs, which is of vital importance for them, as the piglets have not gained mature immune ability. *Lactoferrin (LTF)*, one of the immuno-active materials in milk, possesses antibacterial and antiviral abilities. However, there is a notable reduction of lactoferrin in sow milk during lactation in the first few days, and many piglets could failure to suck enough colostrum under this situation which may increase the morbidity and mortality of piglets. In this study, we successfully constructed genome-edited Large-White pigs with marker-free site-specific knock-in of lactoferrin gene in the 3'-end of *Casein alpha-s1 (CSN1S1)* locus via CRISPR/Cas9 mediated homologous recombination. Thus, the *LTF* protein can be expressed in the mammary gland in the control of *CSN1S1* promoter. As expected, the lactoferrin protein in transgenic pigs sustained high expression in both colostrum and milk when compared with wild type pigs. Moreover, the bacterial plate assay indicated that the milk from gene-targeted pigs showed bacteriostatic effects when compared with control pigs. Taken together, our study demonstrated that the milk from gene-targeted pigs had antibacterial activity which may reduce the costs of veterinary drug and improve the surviving rate of piglets, which is promising for pig breeding.

PO0017: Gene Editing/CRISPR

Bovine Toll-like Receptor 4 Modulates the Inflammatory Response of Mammary Epithelial Cells to *Mycobacterium avium* Subsp. Paratuberculosis Cell Lysate and *Escherichia coli* Lipopolysaccharide

Umesh K. Shandilya¹, Sanjay Mallikarjunappa², Ankita Sharma³, Jiahe Guo⁴, Kristen Lamers³, Mao Yongjiang⁵, Kieran G. Meade⁶ and Niel A Karrow³, (1)Centre for Genetic Improvement of Livestock, Department of Animal Biosciences, Guelph, ON, Canada, (2)University of Guelph, Guelph, ON, Canada, (3)Centre for Genetic Improvement of Livestock, Department of Animal Biosciences, University of Guelph, Guelph, ON, Canada, (4)Yangzhou University, Yangzhou, ON, China, (5)Yangzhou University, Yangzhou, China, (6)Teagasc Animal and Bioscience Research Department, Dunsany, Ireland

Toll-like receptor 4 (TLR4) is a pattern-recognition receptor (PRR) involved in the recognition of microbial pathogens and host alarmins. Ligation to TLR4 initiates a signalling cascade that leads to inflammation. Polymorphisms in bovine TLR4 have been associated with *Mycobacterium avium* subsp. paratuberculosis infection, the cause of Johne's disease, and milk somatic cell score, a biomarker of mastitis. Although the contribution of TLR4 to recognition of bacterial lipopolysaccharide (LPS) has been well characterized, its role in MAP recognition is less certain. Gene editing was performed to generate TLR4 knockout (KO) mammary epithelial (MAC-T) cells to determine if TLR4 expression is involved in the initiation of the host inflammatory response to MAP cell wall lysate (5 and 10 µg/ml) and *Escherichia coli* LPS (5 µg/ml). The expression of inflammatory genes (TNF-α, IL-1α, IL-1β and IL-6) in the LPS-challenged KO cells decreased significantly as compared to unedited cells. However, in response to the MAP cell lysate, the expression of inflammatory genes was higher in TLR4 KO cells than unedited cells. Similarly, higher levels of cytokines/chemokines (TNF-α, IL-6, IL-8, CCL2) were detected in the culture supernatant of TLR4 KO cells when stimulated with LPS, and lower levels were detected when stimulated with MAP cell lysate (TNF-α, IL-6, CCL3, IL-10). Overall, these results indicate that TLR4 is essential for eliciting inflammation in response to LPS; however, exacerbated TLR4 KO cell gene and protein expression in response to MAP cell lysate suggests that TLR4 may be involved in modulating immune signalling.

PE0018: Gene Editing/CRISPR

Tools for Studying Plant Reproductive Biology: Multiplex Gene Editing and High-Throughput Fcsc

Xingliang Ma¹, Rick Goertzen² and Timothy Sharbel¹, (1)Global Institute for Food Security, Saskatoon, SK, Canada, (2)Global Institute for Food Security, Canada

Apomixis is a naturally occurring mechanism of producing seed through asexual reproduction whereby embryos are genetic clones of the mother plant. Apomixis has an enormous potential to improve and accelerate crop production by (for example) fixing complex genetic traits across generations. To investigate this complex mechanism, we use a multiplex CRISPR-Cas9 based gene editing system for validating candidate apomictic genes identified from our apomixis model system (genus *Boechea*, a wild Brassicaceae). These gene edited populations are then

phenotypically analyzed via high-throughput flow cytometric seed screen (hi-FCSS), which allows us to discriminate sexual from apomictic seed lines by precisely determining their embryo and endosperm genome content ratios. We emphasize that the hi-FCSS procedure developed in our lab is also being applied in analyzing reproductive variation in a variety of plant species, including important crops such as canola, lentils, corn, and rice. Our toolbox of gene editing and hi-FCSS plays a critical role for dissecting the genetic factors underlying apomixis in addition to identifying important plant reproductive traits.

PO0019: Gene Editing/CRISPR

Validation of a Gene Controlling Genetic Sexing Trait using CRISPR

Sheina Sim and Scott Geib, USDA-ARS, Hilo, HI

Tephritid fruit flies are a destructive insect family of agricultural pests and are the target of expensive population eradication and suppression efforts within state and federal departments of agriculture in the United States and world-wide. Area-wide integrated pest management programs control species such as the Mediterranean fruit fly and the Mexican fruit fly through the release of sterile males known as the sterile insect technique (SIT). The efficacy of SIT programs is improved through the use of a genetic sexing strain (GSS) which allows for the separation of males from females prior to adult eclosion. To create genetic sexing strains of other high-risk tephritid pests, we performed QTL mapping, differential expression analysis, and comparative whole-genome resequencing to identify candidate genes which control the genetic sexing phenotype for pupal color. Gene function was validated using CRISPR-Cas9 targeted mutagenesis to create knockout mutations via non-homologous end join repair. These methods, coupled with additional non-transgenic techniques for creating reciprocal translocations and fixing sex-linkage, can be applied to other systems to create non-transgenic and science policy friendly genetic sexing strains to control fruit flies as a preventative measure and also future incursions and outbreaks.

PE0020: Gene Editing/CRISPR

Simple Protoplast Regeneration Protocols for Genome Editing

Choun-Sea Lin, ABRC, Academia Sinica, Taipei, Taiwan

Precision genome editing accelerates plant research and crop breeding by modifying the endogenous genes or introducing novel genes to obtain desirable phenotypes. The combination of protoplast transfection and regeneration is a highly efficient approach for DNA-free CRISPR-Cas genome editing and for alleviating issues associated with transgenic plants. However, this combined approach has so far been successful in only limited species owing to the difficulties in protoplast regeneration. Here, we have provided a simple protoplast regeneration protocol that can be easily established in plant research laboratories and readily adopted for genome editing in other model crop species. Genome editing in protoplasts offers an attractive strategy to expedite plant research and breeding, to help study environmental impacts on plants, and to identify solutions for future food security.

PO0021: Gene Editing/CRISPR

Inari : Seeds for a Changing Planet

Catherine Feuillet, Inari Ag, Cambridge, MA

Inari is creating seeds for a changing planet. By combining gene editing technologies with computational science, Inari is dramatically shortening the time and cost it takes to develop a new seed. Using its Seed Foundry concept, Inari is using a variety of cell-based assays and custom-built software to develop a deep understanding of sequence polymorphisms, specific gene networks and quantitative traits. Once the target sequences genes have been identified, Inari generates new allelic diversity using our genome editing technologies and delivers the changes into its elite parental lines. The company's goal is to develop a new generation of more resilient seeds that require less water and use fewer fertilizers and pesticides. Founded by Flagship Pioneering in 2016, Inari is headquartered in Cambridge, Massachusetts with additional sites in West Lafayette, Indiana and Ghent, Belgium. The company was honored as a 2019 Technology Pioneer by the World Economic Forum. To learn more, visit inari.com or follow Inari on Twitter @inari_ag.

PE0022: Gene Editing/CRISPR

Genome Editing in Grass Plants

Bing Yang, University of Missouri, Columbia, MO

Programmable nucleases (e.g., CRISPR RNA guided Cas nucleases) have been successfully engineered to induce site-specific mutations at genomic loci in grass plants such as rice, maize, wheat, sorghum, etc. The genome editing tools have significantly advance our basic understanding of gene function and engineering beneficial traits in grass plants. In my presentation, I will provide our experience in developing and utilizing CRISPR/Cas9 technologies for targeted mutagenesis in several important grass crops.

PO0023: Gene Editing/CRISPR

Overexpression of the Candidate *Pi-ta2* Gene in Rice (*Oryza sativa* L. *japonica* cv. Katy) for Resistance Against Blast Fungus (*Magnaporthe oryzae*)

Gaurav Thapa Chhetri¹, Sathish Kumar Ponniah¹, Yulin Jia² and Muthusamy Manoharan¹,
(1)University of Arkansas at Pine Bluff, Pine Bluff, AR, (2)USDA-ARS, Stuttgart, AR

Rice blast, caused by *Magnaporthe oryzae* is one of the most destructive diseases in rice, which feeds one-half of the world's population. An innate defense mechanism in plants known as effector triggered immunity (ETI) is mediated in rice by blast-resistance (*R*) genes most of which encode cytoplasmic proteins with nucleotide binding site-leucine-rich repeat (NLR) responsible for interacting with fungal effectors to trigger ETI. In the United States, breeding programs utilize a *R* gene cluster presumably carrying *Pi-ta*, *Pi-ta2*, and *Ptr* from the tropical japonica cultivar Katy to confer blast resistance. The objective of this study was to overexpress the candidate *Pi-ta2* gene encoding a NLR protein (chromosome 12; LOC_Os12g18374) located in between *Pi-ta* and *Ptr*, and understand its role in blast resistance in rice. Plant transformation was carried out using 14 day old embryogenic calli cultured on N6D medium containing 3.0mg/L 2, 4-D. Calli were infected with *Agrobacterium tumefaciens* strain EHA105 carrying the binary plant expression vector pCambia 1304 containing the candidate *Pi-ta2* gene. A total of 66 hygromycin resistant calli, obtained from 174 infected ones, were placed on regeneration media containing 30mg/L hygromycin and 200mg/L timentin. Polymerase chain reaction (PCR) was carried out on twelve independent events regenerated from the rooting media using gene specific and hygromycin primers. The results confirmed the presence of respective genes in six plants of T₀ generation. Expression analysis using quantitative real-time transcriptase PCR (qRT-PCR) and scoring after inoculation with blast fungus IB-49 (ML1) will be performed on T₁ plants. Results will help to elucidate the molecular basis of evolutionary mechanisms of rice blast disease *R* genes.

PE0024: Gene Editing/CRISPR

Genome Wide CRISPR Knockout Screen Identifies Host Factors Important for Bovine Herpes Virus Type 1 Replication

Wenfang Spring Tan¹, Enguang Rong², Inga Dry¹, Simon Lillico², Andy Law² and Bruce Whitelaw³,
(1)the Roslin Institute, Midlothian, United Kingdom, (2)the Roslin Institute, United Kingdom, (3)The Roslin Institute, Edinburgh, United Kingdom

CRISPR/Cas9 are molecular scissors that cut DNA in a site-specific manner; it relies on base pairing between the small CRISPR guide RNA (gRNA) and target DNA. This activity leads the gRNA-bound Cas9 protein to the target and exert cutting, creating double strand breaks in the DNA that are repaired by the cell, often resulting in gene inactivating *indels*. Since the specificity of CRISPR/Cas9 is largely determined by the base pairing, it is straightforward to introduce many guides with altered gRNA sequences to achieve knockout of many genes in parallel. To enable whole genome screening in cattle, we utilized this high throughput strategy and produced a CRISPR library containing a pool of 96,000 guides targeting 21,165 protein coding genes in the cow genome. We observed good performance after transducing this library, one guide per cell, into our MBDK cell line that stably expresses Cas9 from the *rosa26* locus: guides targeting core essential genes are significantly depleted while those targeting non-essential genes or non-cutting control guides remain largely unchanged.

Bovine Herpes Virus Type 1 (BHV-1) causes infectious bovine rhinotracheitis, fatalities in calves and pregnancy abortions in cows, leading to huge economy loss to cattle farmers in Ireland and the UK. Unfortunately, little is

known about how host cell factors intervene or facilitate BHV-1 infection, and this lack of knowledge impedes vaccine and drug developments. Thus, to study interactions between BHV-1 and the host, we infected the library transduced cells with a GFP tagged BHV-1 at high MOI, and FACS sorted live cells at 8 hours post infection into sub-populations with different intensities of GFP signals. We identified lists of genes with significantly depleted or enriched guides from these sub-populations, the GFP negative cells in particular. In GFP negative cells, genes targeted by enriched guides are candidates that facilitate or are essential for BHV-1 infection whereas genes targeted by depleted guides may inhibit this virus. We are currently validating candidate genes using a focused library that targets only the candidates from the genome wide screen. We are also using drug inhibitors and individual gRNAs that knockout, activate or repress gene expressions to study their detailed mechanisms.

PO0025: Gene Editing/CRISPR

Transgene Stacking in Potato Using the Gaantry System

James Thomson¹, Kent F. McCue², Ethan Gardner³, Ron Chan³ and Roger Thilmony¹, (1)USDA-Agriculture Research Service, Albany, CA, (2)USDA-Agricultural Research Service, Albany, CA, (3)USDA-Agriculture Research Service

This research presents a novel plant biotechnology for the rapid genetic improvement of crops. Although single genes have been important in engineering herbicide and pest tolerance traits, future improvements of complex traits like yield and nutritional quality will likely require the introduction of multiple genes. The GA4NTRY (Gene Assembly in Agrobacterium by Nucleic acid Transfer using Recombinase technology) system as a flexible and effective system for stably stacking multiple genes within an *Agrobacterium* virulence plasmid Transfer-DNA (T-DNA). The system utilizes unidirectional site-specific recombinases *in vivo* and an alternating selection scheme to sequentially assemble multiple genes into a single transformation construct. To demonstrate GA4NTRY's capabilities, 10 cargo sequences were sequentially stacked together to produce a 28.5 kilobase pair T-DNA, which was used to generate transgenic potato. Approximately 89% of the events identified using the dual antibiotic selection screen exhibited all of the introduced traits. A total of 57% of the tested lines carried a single copy of the selection marker transgene located near the T-DNA left border and none of the plant tested contained sequence from outside the T-DNA. These results demonstrate that the GA4NTRY is a powerful, yet simple to use, new tool for transgene stacking, plant synthetic biology and the generation of high quality genetically engineered plants.

PE0026: Bioenergy

Collaborating on Data, Science, and Infrastructure to Advance Computational Systems Biology Research in KBase

Vivek Kumar¹, Sunita Kumari², Doreen Ware², Christopher Henry³, Robert W. Cottingham⁴, Shane Canon⁵, Paramvir Dehal⁵, Roy Kamimura⁵, Elisha Wood-Charlson⁵ and Adam Arkin⁵, (1)Cold Spring Harbor Laboratory, NY, NY, (2)Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, (3)Argonne National Laboratory, Argonne, IL, (4)Oak Ridge National Laboratory, Oak Ridge, TN, (5)Lawrence Berkeley National Laboratory, Berkeley, CA

The Systems Biology Knowledgebase (KBase) is an open-science, open-source, software and data platform funded by U.S. Department of Energy in support of biological and environmental research to investigate the complex interactions within biological systems. Supported by scalable computing infrastructure, it enables the sharing, integration, and analysis of multi-omic datasets related to prokaryotes and eukaryotes, particularly microbes, plants, fungi and their communities. Thus, it lowers the barrier to accessing computational tools, data, and results, and to work collaboratively to accelerate the pace of research.

KBase offers a suite of scientific applications to enable users to build sophisticated analytical workflows and share their findings. The nearly 200 apps in KBase offer diverse scientific functionality across the realms of comparative genomics, community analysis, metabolic modeling, and transcriptomics. For example, one could predict species interactions from metagenomic data by assembling raw reads, binning assembled contigs by species, annotating genomes, aligning RNA-seq reads, and reconstructing and analyzing individual and community metabolic models.

Users can build and share sophisticated workflows through a combination of chaining together multiple analysis tools, writing scripts for automation, and using batch processing, all within notebook-style Narratives that contain the employed data and tools. Developers can build, test, register, and deploy new or existing software as KBase apps using the Software Development Kit, thereby extending the platform's scientific capabilities.

Recently developed features allow for greater organization of collaborative projects and increased depth of discovery within massive datasets. Projects, laboratories, and even whole institutions can organize their users and associated Narratives into a shared Organization with multiple permission levels and management features. Additionally, an early version of social feeds informs users of changes happening within their Organizations. Newly added services enable the platform to find and suggest data sets or Narratives that may be of interest to a particular user, based on searching interconnections between the data in KBase. KBase is unique in offering these diverse and integrated capabilities to a growing user community that is actively pioneering the use of the platform in their publications.

KBase is funded by the Genomic Science program within the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research under award numbers DE-AC02-05CH11231, DE-AC02-06CH11357, DE-AC05-00OR22725, and DE-AC02-98CH10886.

PO0027: Bioenergy

GWAS Analysis Reveals Genetic Factor Controlling Bud Break in *Populus trichocarpa*

Kai Feng, Oak Ridge National Laboratory, Oak Ridge, TN

Many perennial plants become dormant to survive freezing winter and dehydration, when bud resumes to growth known as bud break, is a sign of end of dormancy. The activity-dormancy cycle is crucial for both survival and growth of plants. To understand bud dormancy, phenotyping of bud break were applied in three common gardens of 1,146 *Populus trichocarpa* genotypes across multiple years. We then performed genome-wide association studies (GWAS) incorporating bud break phenotypic data with 8,301,860 SNPs and InDels (minor allele frequency > 0.05) from 917 *P. trichocarpa* accessions and expression-based quantitative trait loci (eQTL) analyses to identify key regulators. All bud break phenotypes are significant correlated with an InDel in the promoter region of a UDP-sugar transporter protein (PtUTr) after Bonferroni correction. This InDel is found to be an AT-Hook binding domain which often serves as a cis-acting element in plants. PtUTr leaf and xylem expression were significantly correlated with bud flush phenotypic data and it was regulated by cis-eQTLs containing the AT-Hook binding domain. This study provides insights into data-driven of gene function in woody species.

PE0028: Bioenergy

The Genetic Relationship and Potential Heterotic Groups within a Diverse Set of Cytoplasmic Male Sterile Lines of Sorghum

Smit Dhakal¹, Lynn Doran¹, Marlee Rose Labroo¹, Patrick J. Brown² and Anthony J. Studer¹,
(1)University of Illinois at Urbana-Champaign, Urbana, IL, (2)Department of Plant Sciences, University of California, Davis, Davis, CA

The United States Environmental Protection Agency mandates production of 21 billion gallons of alternative fuels from non-cornstarch based sucrose. Ethanol production in the United States is approximately 16 billion gallons as of 2018. Thus, a renewable fuel source is needed to fill this gap in production and contribute to energy security in the future. Sorghum (*Sorghum bicolor*) is a promising crop for bioenergy production due to its high yield potential and wide adaptability. In addition, sorghum has extensive natural genetic variation that can be leveraged for breeding favorable traits for bioenergy production. However, due to the sorghum's hermaphroditic panicle, crosses between lines is extremely difficult. The use of cytoplasmic male sterility is one method to overcome this challenge to crossing, but there is a limited amount of diversity present in the female lines used for hybrid production. The objective of this study was to develop a genetically diverse set of cytoplasmic male sterile lines that can be used as females for testing hybrid combinations. To accomplish this we backcrossed cytoplasmic male sterility for five generation into 30 sorghum conversion lines. These lines were genotyped with 100K (genotype by sequencing) markers to assess relatedness, and to test for a link between genetic diversity and heterosis. These lines were then

used to make hybrids, and data was collected for stem sugar content, juice volume, stem composition, biomass and seed yield were taken. The results of these analyses will allow us to better select complementary lines that will produce superior bioenergy sorghum hybrids.

PO0029: Bioenergy

Nutritional Composition of Commercial Sugarcane (*Saccharum* spp.) Genotypes Evaluated in Different Environments of Brazil

Monalisa S. Carneiro¹, Rodrigo Gazaffi¹, Nara F. Pistarini¹, Rodolpho T. G. Lorencetti¹, Eduardo A. Bressan², Edson F. Silva³, Giancarlo C.X. Oliveira⁴ and Maria Teresa M.R. Borges¹, (1)Federal Univ. of São Carlos, Araras, Brazil, (2)Dep. of Genetics, Luiz de Queiroz College of Agriculture, Univ of São Paulo, Piracicaba, Brazil, (3)Dep. of Biology, Federal Rural University of Pernambuco, Recife, Brazil, (4)Dep. of Genetics, Luiz de Queiroz College of Agriculture, Univ. of São Paulo, Piracicaba, Brazil

Sugarcane is one of the most economically important grasses in the world. This research focuses on assessing genetic variation of sugarcane nutritional composition (NC) traits, and investigates genotype-by-environment interactions. Twenty relevant sugarcane commercial cultivars were evaluated for 20 NC traits over six environments in the main productive regions of Brazil during two years (first and second ratoons). A high variability of the NC traits associated to the sugar content was found among the cultivars, which ranged from 15.8 to 23.1°(Brix); 12.8 to 19.8% (Pol of Juice); 10.2 to 15.9% (Pol of Cane by Tanimoto's method); 10.4 to 16.4% (Pol of Cane by the PCTS [*Payment for Sugarcane by Sucrose Content*] method); 117.8 to 175.6 kg/tonnes (Total Recoverable Sugars); 11.6 to 17.8% (Total Recoverable Sugars of Cane); 11.5 to 17.3% (Total Recoverable Sugars of Cane by Tanimoto's method). Also, a high amplitude was observed in fiber (10.8 to 18.4%), fiber by Tanimoto's method (9.1 to 19.6%), purity (75.4 to 89.9%) and humidity (64.3 to 76.0%). The correlations between the traits associated with sugar content were high; however, those between fiber and sugar traits were low. The data will be incorporated in a sugarcane NC public database. Information on the NC of sugarcane will be used as a reference in biosafety evaluations (namely, substantial equivalence) of GM cultivars.

Financial support: FAPESP (2013/50195-2)

PE0030: Bioenergy

Isoform-Aware Expression Analyses in Sugarcane Enhance the Identification of Candidate Genes for Fiber Content

Gabriel R A Margarido¹, Fernando H. Correr¹, Guilherme K. Hosaka¹, Fernanda Z. Barreto², Thiago W. A. Balsalobre² and M.S. Carneiro³, (1)University of Sao Paulo, Piracicaba, Brazil, (2)Federal University of Sao Carlos, Araras, Brazil, (3)Dep. of Biotechnology and Plant and Animal Production, Center for Agricultural Sciences, Federal University of São Carlos, Araras,, Brazil

The genus *Saccharum* includes complex polyploid species and interspecific hybrids that show high potential for sugar and biomass production. These genotypes have large, repetitive and highly heterozygous genomes, with multiple polyploidization events, which cripple advances in the application of functional genomics. Most RNA sequencing assays in sugarcane do not leverage information about gene isoforms, thus providing a limited overview of the transcriptional landscape. In this context, little is known about the effects of the various levels of gene expression control on phenotypic outcomes in these highly polyploid genomes. In this study we evaluate *Saccharum* genotypes contrasting in fiber accumulation and show that analysis at the transcript level provides complementary information about differential gene expression. We used RNA-seq to evaluate the expression profiles of leaves from 12 sugarcane genotypes with low or high fiber content. For 5886 genes we found evidence of differential expression at the gene level (DEG), with no significant differences for their individual isoforms. Conversely, 8693 genes revealed at least one differentially expressed transcript (DET), even though their gene-level abundances were similar for both fiber groups. As a consequence, for instance, we noted that photosynthetic processes were not enriched among DEGs, but were significantly overrepresented among DETs. This observation shows that assessing both gene- and transcript-level quantifications offers a more detailed view of the expression status of polyploid sugarcane. Simply aggregating isoforms to study gene expression may bias the list of identified candidate genes, as well as the results of functional enrichment analyses.

PO0031: Bioenergy

Sequence Polymorphism of Phytochrome B Homologs on Flowering Time in *Miscanthus*

Zhihui Guo¹, Hironori Nagano², Lindsay V. Clark³, Erik J. Sacks³ and Toshihiko Yamada², (1)Graduate School of Environmental Science, Hokkaido University, Sapporo, Japan, (2)Field Science Center for Northern Biosphere, Hokkaido University, Sapporo, Japan, (3)Dept of Crop Sciences, University of Illinois at Urbana-Champaign, Urbana, IL

The perennial C₄ grass *Miscanthus* is native to eastern Asia and Oceania with high yields and less fertilizer and herbicide inputs, which is suggested as one of the most promising bioenergy crops. Flowering time has a significant impact on plant adaptation to agro-ecological environments, biomass accumulation and grain yield, and it is regulated by plant development, photoperiod, shading, temperature, nutrient status, and many other factors. Phytochrome B (*PHYB*) is a photosensory receptor and identified as one of key factors regulating photoperiod - dependent flowering in plant including rice, barley and sorghum. The previous study revealed that overexpressing *Arabidopsis PHYB* in transgenic *M. sinensis* is functional that phenotypes with increased chlorophyll content, decreased plant height, and delayed flowering, but little is known about characterization of *PHYB* in *Miscanthus*. Here we investigated the whole genomic sequences of *PHYB* in twelve *Miscanthus* accessions originated from Asian mainland and Japanese archipelago, representing different latitudes. As a whole genome-duplication existed in *Miscanthus* relative to *Sorghum bicolor*, two *MsiPHYB* homologs were identified with high structural similarity. The conserved domains of *PHYB*, such as GAF domain, Phytochrome region, PAS domain and Histidine kinase domain, were also observed in the predicted protein sequence of *MsiPHYB*. Relation of the allelic variation and distribution across region and latitude of wild *Miscanthus* were observed. The finding is useful for the understanding of genetic control of flowering time in *Miscanthus*, and may facilitate to further develop varieties for specific environments with enhanced agricultural performance using phytochromes.

PE0032: Other Category

Addressing Severe Food Shortages in Africa—the Role of the Makerere University Regional Centre for Crop Improvement

Paul T Gibson¹, Isaac Dramadri², Patrick Obia Ongom^{2,3} and Richard Edema¹, (1)Makerere University Regional Centre for Crop Improvement (MaRCCI), Kampala, Uganda, (2)Makerere Univ. Regional Centre for Crop Improvement (MaRCCI), Kampala, Uganda, (3)IITA (International Institute of Tropical Agriculture), Kano, Nigeria

A dramatic increase in agricultural production in Sub-Sahara Africa (SSA) is needed to avoid severe food shortages within a few years. Such increase can happen only with the convergence of relevant government policies, strategic investment, marketing arrangements, production management, and improved breeds of both animals and plants. Capable, innovative and effective professionals from all disciplines are still too few in SSA to address these needs. Makerere University Regional Centre for Crop Improvement (MaRCCI), based in Kampala, Uganda, is meeting part of that need by preparing well-trained plant breeders from over 20 African countries in up-to date practices, focusing on the many varied crops that are important for food, nutrition, and income security in SSA.

In cooperation with the College of Agriculture's Department of Agricultural Production, MaRCCI offers an MSc in Plant Breeding and Seed Systems, and a PhD in Plant Breeding and Biotechnology, as well as periodic training workshops focused on skills needed for modernizing plant breeding and variety delivery. The training emphases for all these include -- development of variety descriptions that meet producer, consumer and value-chain needs; optimization of the breeding pipeline; e-capture of performance data; database management; and high-throughput genotyping and phenotyping. Instruction incorporates approaches that are learner-centered and E-based, including material from the Plant Breeding E-learning in Africa (PBEA) program. "Best-practice" in-house breeding programs target climate-resilient crops (cowpea and sorghum), providing practical application of the scientific principles emphasized in classwork. Thesis research is embedded in national and international research programs (eg. NARS & CGIAR's) in order to address critical bottlenecks in variety improvement and to impact delivery in the region. Priorities for training activities use Graduate Profiles and Core Competencies to produce "fit-for-purpose" "market-ready" graduates. Regional and global partnerships with premier educational and research organizations and seed

companies enhance the students' learning opportunities. A globally prestigious Advisory Board includes representatives of diverse stakeholders.

Graduates, employers, and development partners have often expressed strong approval of the program. World Bank has designated MaRCCI as an African Centre of Excellence in Graduate Plant Breeding training and research. Continuous external and internal evaluation promotes excellence and relevance. Cost is reduced by in-region training. Careful student selection, geographical proximity of the program to the students' home area and gender-sensitive provisions have resulted in almost 100% completion and in-region retention of the more than 140 MSc and 70 PhD students who have completed or are in-process over the past 11 years.

PO0033: Other Category

National Center for Biotechnology's (NCBI) New Search Functionality for Genes and Genomes

Nuala A. O'Leary, Terence D. Murphy, Valerie A. Schneider and Peter Meric, NCBI/NLM/NIH, Bethesda, MD

NCBI recently made available a new and improved search experience that interprets plain language for commonly performed queries within the sequence databases including Nucleotide, Protein, Gene, Genome, Assembly, and the "All Databases" search page. The new search functionality addresses well-defined queries that previously failed to return relevant results, such as organism-gene (e.g. sheep LEP) or organism-assembly (e.g. tomato reference genome). Results are presented in new, easy-to-interpret interfaces at the top of results pages and highlight the data and relevant tools likely to be of greatest interest to most users. To further facilitate searches, as-you-type suggestions have been added to the search bar in these databases. We have also added a new way for users to find evolutionarily related genes within and across organisms represented in the NCBI RefSeq dataset. This poster will summarize this new NCBI search functionality and present our current efforts to allow users to download customized datasets.

PE0034: Other Category

Identifying and Addressing Naïve Concepts in the Undergraduate Genetics Student

Rebecca L. Seipelt-Thiemann, Middle Tennessee State University, Murfreesboro, TN

Genetics is known to have difficult concepts to learn and understand due to its use of complex, nuanced, and technical vocabulary, as well as the abstract nature of many concepts, the requirement to think across multiple scales, particularly the symbolic scale, and the pace of change to the field. Concept inventories are assessment tools that can be focused on a particular concept or a course and can aid faculty identify and address student naïve concepts as well as be used as assessment tools for establishing whether particular learning activities are helpful to student learning. In the workflow used in the development and validation of four genetics-focused concept inventories, 1) mutation, 2) pedigree analysis, 3) epistasis, and 4) epigenetics, common student errors and broader misconceptions common to several, and sometimes all, concept areas were identified. Student responses showed that students have major issues understanding: 1) the difference between a mutation and any change in RNA, protein, or function, 2) specialized terminology, 3) the flow of genetic information in both molecular and inheritance aspects, and 4) cross-scale relationships, such as mutation to heritability, gene to chromosome, gene to allele, and mathematical probability to phenotype. It is hoped that knowing which alternative conceptions students commonly hold will aid faculty in designing instruction that enables students to form a more accurate conceptual framework regarding genetics concepts.

PO0035: Other Category

Phytobiomes Research for Enhancing the Sustainable Production of Food, Feed, and Fiber

International Alliance for Phytobiomes Research, International Alliance for Phytobiomes Research, Lee's Summit, MO, Gwyn A. Beattie, Iowa State University, Ames, IA, Natalie W. Breakfield, NewLeaf Symbiotics, Saint Louis, MO, **Kellye Eversole**, International Phytobiomes Alliance, Lee's Summit, MO, Magalie Guilhabert, Bayer CropScience, Biologics, West Sacramento, CA, Jan E. Leach, Colorado State

University, Fort Collins, CO, Matthew J. Ryan, CABI, Egham, Surrey, United Kingdom and Angela Sessitsch, AIT - Austrian Institute of Technology, Vienna, Austria

A major paradigm shift in agricultural production is required to meet the demands of a global world population projected to reach 9.7 billion in 2050. We need to sustainably increase crop productivity, while preserving biodiversity, natural resources, and grower income in the context of climate change. To optimize sustainable productivity and profitability on farms, grasslands, and forests, scientists must embrace a holistic, systems-level approach and focus on the complexity within phytobiomes. The term “phytobiome” refers to a plant growing within a specific environment, or biome, and all of the micro- and macro-organisms living in, on, or around it—such as microbes, animals, insects, and other plants—as well as the geophysical environment, which includes soil, air, water, weather, and climate. By establishing a foundation of knowledge on how phytobiome components interact and affect each other, the Phytobiomes Alliance (www.phytobiomesalliance.org) a non-profit alliance of industry, academic, and governmental partners created in 2016, aims at addressing today’s agricultural challenges. The Alliance facilitates and coordinates international efforts toward expanding phytobiomes research in order to accelerate the sustainable production of food, feed, and fiber for food security. Current priority areas of the Alliance include filling the gaps in our knowledge of how microbes interact with other phytobiome components in outdoor and controlled environments as well as building a regulatory science foundation to support rapid commercialization of sustainable, microbial based products that increase the productivity and viability of agricultural production systems.

PE0036: Other Category

Enabling Breeders to Access Modern Genotyping Tools and Services in Africa for Genetic Gain Enhancement: CGIAR Excellence in Breeding Platform Services

Ana Luisa Garcia-Oliveira¹, Eng Hwa Ng², Rajaguru Bohar³, Manilal William⁴, Milcah Kigoni⁵, Rajeev K Varshney³, Petra Vanroggen⁶, Patrik Stolt⁷, Jean-Marcel Ribaut⁸, Michael Quinn⁹ and Michael Olsen¹⁰, (1)EiB, Ibadan, Nigeria, (2)Excellence in Breeding, Laguna, Philippines, (3)ICRISAT, Hyderabad, India, (4)Integrated Breeding Platform, Texcoco, VL, Mexico, (5)Intertek AgriTech, Alnarp, Sweden, (6)Intertek ScanBi Diagnostics AB, Alnarp, Sweden, (7)Intertek, Alnarp, Sweden, (8)Integrated Breeding Platform, Texcoco, Mexico, (9)Excellence in Breeding, CIMMYT, Texcoco, Mexico, (10)CIMMYT, Nairobi, Kenya

The CGIAR Excellence in Breeding (EiB) platform is aimed to empower and support the streamlining of breeding programs among CGIAR centers and national agricultural research system partners in developing countries across the continents including Africa. Breeding program members are supported by EiB in various ways including the mainstreaming of genotypic data utilization to improve breeding program efficiency. This is done facilitating access to user-validated and application-appropriate genotyping tools. These actions are important in Africa, particularly in the Sub-Saharan region, where diverse breeding teams require access to cost effective genotyping services with logistical support and rapid data return to implementation of marker-assisted breeding flows. Although modern genotyping methods are routine and high throughput, still there remains a need to implement streamlined processes to improve efficiency, reduce error, and achieve lowest pricing options with vendors. This includes sample collection, laboratory scheduling and forecasting, and high-quality data delivery and analysis to support breeding programs at a lower cost. In partnership with the High throughput genotyping services (HTPG), INTERTEK genotyping service provider and the Integrated Breeding Platform (IBP), establishing protocols and user documents, mediating genotyping services with vendors, and delivering high quality genotypic data that can be readily interpreted and utilized using user-friendly tools. The ultimate goal is to accelerate genetic gains in the diverse breeding programs of the region through efficient and appropriate application of genotypic data.

Key words: EiB, genotyping, plant crops, genetic gains, molecular markers, streamlining tools

PO0037: Other Category

Deep Active Learning Framework for Image-Based Plant Phenotyping

Koushik Nagasubramanian¹, Talukder Zaki Jubery², Seyed Vahid Mirnezami³, Asheesh K. Singh², Arti Singh², Soumik Sarkar³ and Baskar Ganapathysubramanian³, (1)Department of Electrical Engineering,

Iowa State University, Ames, IA, (2)Department of Agronomy, Iowa State University, Ames, IA, (3)Department of Mechanical Engineering, Iowa State University, Ames, IA

Deep learning models have been highly successful in the image-based plant phenotyping applications such as disease detection and classification. However, one of the main challenges in achieving this success is the requirement of large amount of labeled data. Data annotation could be costly, time consuming and hard for many plant phenotyping tasks. To overcome this challenge, recently many active learning algorithms have been proposed to reduce the amount of labeling needed by deep learning models for achieving high performance. We propose a task agnostic active learning framework for a deep learning model to achieve maximal performance under a fixed labeling budget. We evaluate the performance of our active learning framework for different image-based plant phenotyping applications.

PE0038: Methods: Cellular Processes and Regulatory Networks

Comparing Time Series Transcriptome Data between Plants Using a Network Module Finding Algorithm

Jiyoung Lee, Ruth Grene, Lenwood S. Heath and Song Li, Virginia Tech, Blacksburg, VA

Background

Comparative transcriptome analysis is the comparison of expression patterns between homologous genes in different species. Since most molecular mechanistic studies in plants have been performed in model species, including Arabidopsis and rice, comparative transcriptome analysis is particularly important for functional annotation of genes in diverse plant species. Many biological processes, such as embryo development, are highly conserved between different plant species. The challenge is to establish one-to-one mapping of the developmental stages between two species.

Results

In this manuscript, we solve this problem by converting the gene expression patterns into co-expression networks and then apply network module finding algorithms to the cross-species co-expression network. We describe how such analyses are carried out using bash scripts for preliminary data processing followed by using the R programming language for module finding with a simulated annealing method. We also provide instructions on how to visualize the resulting co-expression networks across species.

Conclusions

We provide a comprehensive pipeline from installing software and downloading raw transcriptome data to predicting homologous genes and finding orthologous co-expression networks. From the example provided, we demonstrate the application of our method to reveal functional conservation and divergence of genes in two plant species.

PO0039: Methods: Cellular Processes and Regulatory Networks

Identification of Differentially Expressed Proteins Involved in Plant Resistance to Aphid

Jian Huang, Oklahoma State University, Stillwater, OK and Yinghua Huang, USDA ARS, Stillwater, OK

Greenbug, *Schizaphis graminum* (Rondani), is known as a major insect pest of small grains throughout the world, including wheat, barley and sorghum. Resistance sources in those crops have been identified and are being widely used to manage this important aphid pest; however, little is known about the mechanism of host plant resistance to greenbug. The previous studies demonstrated that resistant plants initiated their innate defense systems in response to greenbug feeding based on the transcriptional profiles developed in sorghum plants. In this study, we undertook proteomic analysis of aphid-challenged sorghum seedlings to compare expression changes at the protein level between greenbug resistant and susceptible lines, PI550610 (resistant) and BTx623 (susceptible) in order to

elucidate molecular mechanisms of insect resistance in sorghum. Seedlings of two sorghum lines were infested with freshly prepared virulent greenbug aphids and leaf tissues were collected on fifth day after infestation. Then total proteins were extracted for the leaf samples and analyzed in parallel using two dimensional gel electrophoresis (2-DE) coupled with mass spectrometry (MS). At least 108 proteins spots reproducibly detected on the protein gels. Of these, comparing resistant and susceptible lines, twenty differentially expressed proteins were identified by MALDI-TOF/MS. Putative function of some differentially expressed proteins is inferred from metabolic processes to plant defense, while others showed as novel genes that need to be confirmed experimentally. We believe that those proteomic data provide new insights on plant defensive responses to greenbug attack in sorghum.

PE0040: Methods: Cellular Processes and Regulatory Networks

Is Epigenetics Involved in the Regulation of Biomineralization in *Emiliania huxleyi*

Christopher J. Smith, California State University San Marcos, San Marcos, CA and Betsy Read, California State University, San Marcos, San Marcos, CA

Emiliania huxleyi (*E. huxleyi*) is one of the most abundant marine phytoplankton in today's oceans. The tiny coccolithophore produces intricate calcium carbonate disks known as coccoliths that surround the cell. In 1983 samples of *E. huxleyi* collected off of the coast of Peru were isolated and sent to two culture collections: 1) the National Center for Marine Algae and Microbiota in Bigelow Maine, and 2) the Plymouth Algal Culture Collection in England. After 20 years in culture, CCMP 1516 has lost, and PLY 217 retained its ability to calcify. As isogenic lines it is hypothesized that epigenetic changes may be responsible for these phenotypic differences. Epigenetics refers to heritable changes in gene expression that do not involve alterations in the DNA sequence. These changes include methylation, phosphorylation, and or ubiquitination of DNA bases and/or histone proteins that influence genome packaging and ultimately gene expression. To determine whether methylation may be an epigenetic driver of calcification in *E. huxleyi*, bisulfite sequencing and RNA-Seq were employed to compare DNA methylation and gene expression levels in CCMP 1516 and PLY 217. Comparisons revealed 12,800 differentially methylated regions and 14,525 differentially expressed genes. Of these, 2,037 genes were differentially expressed and differentially methylated near the transcriptional start site. To independently validate expression, 9 genes previously identified as potentially involved in biomineralization were subjected to Real Time RT-PCR. Differential expression of 5 of the 9 genes tested was confirmed, including that of a Plasma membrane type H pump/ATPase, nucleoside transporter, acyl-coA-binding protein, bacterial Na⁺/H⁺ exchanger, and an unknown protein. Differential expression levels varied from 2-10 . Results from this work will be presented together with preliminary data aimed at profiling histone modifications as an epigenetic mechanism that may be involved in the regulation of calcification and coccolithogenesis.

PO0041: Methods: Cellular Processes and Regulatory Networks

Determining If the K(3) Herbicide Cafenstrole Affects Alkenone Biosynthesis in *Emiliania huxleyi*

Sergio S Serrato-Arroyo, California State University San Marcos, San Marcos, CA and Betsy Read, California State University, San Marcos, San Marcos, CA

Emiliania huxleyi (*E. hux*) is one of the most abundant phytoplankton on this planet. In addition to producing intricate calcium carbonate cell coverings, they are one in five species on the planet to synthesize abnormally long chain fatty acids known as alkenones. Unlike other fatty acids carbon chain lengths range from 36-41. They have one to five *trans* double bonds and exist predominately as free fatty acids opposed to being esterified to a glycerol molecule. Most phytoplankton store energy in the form of triglycerides but on the other hand *E. hux* stores energy as alkenones. For many years' biogeochemists have used the alkenone desaturation index to estimate sea surface temperatures and construct paleoenvironments. The pathways involved in the biosynthesis and degradation of alkenones however, remain unknown. To determine whether elongases are involved in the synthesis of alkenones we used the K(3) herbicide Cafenstrole known to inhibit cell division and synthesis of very-long-chain fatty acids. To this end *E. hux* cells were plated on a lethal dose of Cafenstrole. Two spontaneous mutants were selected and characterized after growing in batch culture. The growth rates were determined, and neutral lipids were extracted and profiled using gas chromatography-mass spectrometry (GC-MS). The growth rates of the mutants and the wild type were similar whereby the doubling time of the wild type strain CCMP 1516 was 31.5 hrs, Caf Sp1 was 32.5 hrs, and Caf Sp2 was somewhat greater at 39 hrs. In terms of total neutral lipids, the two cafenstrole mutants produced

50% less alkenones and nearly 50% more alkanes compared to the wild-type CCMP 1516. RNA was extracted for transcriptional profiling, and genomic DNA was isolated to identify genomic lesions in the mutants. RNA analysis indicates three elongases are down regulated in the mutant strains. The expression profiles of these and several other genes related to lipid metabolism that are significantly differentially expressed will be described, and progress identifying genomic lesions will also be detailed.

PE0042: Methods: Cellular Processes and Regulatory Networks

Quantitative, Super-Resolution Imaging of Small RNAs with sRNA-PAINT

Kun Huang¹, Feray Demirci², Blake Meyers³ and Jeffrey Caplan¹, (1)University of Delaware, Newark, DE, (2)FiDoSoft Software Consulting, Redmond, WA, (3)Donald Danforth Plant Science Center, St. Louis, MO

Small RNAs are 21- to 24-nt non-coding RNAs that play important regulatory roles in the life of both animals and plants. Their small size and high diversity made it challenging to develop methods that have sufficient resolution and specificity. We created a method sRNA-PAINT, for the detection of small RNA with nanometer resolution. Our method utilizes the high-resolution and quantification advances in DNA-PAINT (DNA-based points accumulation in nanoscale topography) methodologies, and combines the specificity of locked nucleic acid (LNA) *in situ* detection of small RNAs. We applied sRNA-PAINT for detecting and quantifying small RNAs in different cell layers of early developmental stage maize anther that are important for male sexual reproduction.

PO0043: Methods: Cytology

myTags[®] - Synthetic FISH Probes Design Schemes for Plant and Animal Molecular Cytogenetics – Recombination, Translocation, Ploidy and Karyotyping Applications

Linda Barthel, Arbor Biosciences, Ann Arbor, MI

Developing the karyotype of most non-model plant and animal species relies on identification of individual chromosomes, which has been a major challenge. We have developed a technology to design, synthesize and label custom target-specific synthetic oligonucleotide fluorescent probes (myTags[®]) for fluorescent *in situ* hybridization (FISH). Through multiple collaborations we demonstrated these probes can be used to uniquely index chromosomes for rapid identification, from both diploid and polyploid species, paint whole chromosome to identify rearrangements such as translocations, or follow haplotypes through meiotic crossovers over multiple generations. These techniques were proven in multiple plant species but could also be applied to animal species. Probes designed from the potato genome were successfully used to identify the 12 homeologous chromosomes among distantly related Solanum species, including tomato and eggplant. In maize, whole chromosome painting demonstrated ability to identify translocations between chromosomes, while haplotype-specific chromosome painting permitted to follow parental and recombinant chromosomes in F1 hybrids and F2 progenies. We believe these techniques can be applied to a wide range of plant and animal species for analyzing recombination, rearrangements, karyotypes and chromosomal relationships for fundamental research and breeding purpose.

PE0044: Methods: Cytology

RNA-Guided Endonuclease – *in situ* Labelling (RGEN-ISL): CRISPR-Based Imaging of Genomic Sequences in Plants and Animals

Takayoshi Ishii¹, Alžběta Němečková², Solmaz Khosravi³, Steven Dreissig³, Bhanu Prakash Potlapalli³, Veit Schubert³, Joerg Fuchs³, Janina Metje⁴ and Andreas Houben³, (1)Arid Land Research Center, Tottori university, Tottori, Japan, (2)Institute of Experimental Botany, Czech Academy of Sciences, Center of the Region Haná for Biotechnological and Agricultural Research, Olomouc, Czech Republic, (3)Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Seeland, Germany, (4)Julius Kühn-Institute, Institute of Biosafety in Plant Biotechnology, Seeland, Germany

We developed a tool to visualize defined genomic sequences in fixed nuclei and chromosomes based on a two-part guide RNA with the recombinant Cas9 endonuclease complex. In contrast to classical *in situ* hybridization, RGEN-ISL (RNA-guided endonuclease – *in situ* labelling) does not require DNA denaturation and therefore permits a

better structural chromatin preservation. The application of differentially labelled tracrRNAs allows the multiplexing of RGEN-ISL. We established a combination of RGEN-ISL, immuno-staining and EdU labelling to visualize *in situ* specific repeats, histone marks and DNA replication sites, respectively. The broad range of adaptability of RGEN-ISL to different temperatures and combinations of methods has the potential to advance the field of chromosome biology.

PO0045: Methods: Functional Analysis

The FAANG Data Coordination Centre: Providing Infrastructure for Functional Annotation of Livestock Genomes

Alexey Sokolov, Jun Fan, Guy Cochrane, Daniel R. Zerbino, Peter W Harrison and Paul Flicek, European Bioinformatics Institute (EMBL-EBI), Cambridge, United Kingdom

To facilitate understanding of genotype to phenotype links in livestock animals, The Functional Annotation of Animal Genomes (FAANG) Project provides the scientific community with high quality functional annotation of livestock genomes. The FAANG Data Coordination Centre (DCC) at EMBL-EBI develops the core infrastructure to support this aim. We focus on ensuring high quality metadata that describes samples, experiments and analysis assays, and to support the community in this we provide an active helpdesk (faang-dcc@ebi.ac.uk) and metadata validation and conversion tools to support submission of data to public archives. FAANG strongly supports open science and has a rapid pre-publication policy for experiment and analysis data (<http://www.faang.org/data-share-principle>). To provide easy access to livestock functional annotation data the DCC has created the FAANG data portal (<https://data.faang.org>), where users can get access to the wealth of livestock annotation data submitted by FAANG contributors and other data from public archives submitted under legacy standards. The portal provides rich filtering and search capabilities, direct links to download data from public archives (including new bulk download support) and programmatic API access that help livestock community to identify data appropriate for their research. It is possible to browse through available sampling and experimental protocols and we are currently incorporating the FAANG validation and submission procedures into the data portal to make it a single access point for all FAANG data access and submission requirements. We also provide a graphical summary of all data currently available in FAANG and perform automated literature scanning to identify data used in published papers. We also continue to evolve the metadata standards to respond to technological and community developments, for example our recent support for CAGE-Seq data. Through these services we help the scientific community to find and establish new links between the genome and phenome that are important for future development of agriculture.

PE0046: Methods: Functional Analysis

Haplotype- and Sequence-Based Identification of a Deletion Associated with Early Embryonic Loss in Holstein Cattle and Functional Validation using CRISPR-Cas9 Knockouts

Derek M. Bickhart¹, **John B. Cole**², Jana L. Hutchison², Daniel J. Null² and M. Sofia Ortega³, (1)Dairy Forage Research Center, USDA-ARS, Madison, WI, (2)Animal Genomics and Improvement Laboratory, USDA-ARS, Beltsville, MD, (3)Division of Animal Sciences, University of Missouri, Columbia, MO

We present a comprehensive framework for identification and validation of genetic defects, including haplotype-based detection of defects, selection of variants from sequence data, and *in vitro* validation using CRISPR-Cas9 knockout embryos. Holstein Haplotype 2 (HH2), which causes early embryonic death, was used to demonstrate the approach. HH2 was identified using a deficiency-of-homozygotes approach and confirmed to have undesirable effects on conception rate and stillbirths. Five carriers were present in a group of 183 sequenced Holstein bulls selected to maximize the coverage of unique haplotypes. Three variants concordant with the haplotype calls were found in HH2: a high-priority frameshift mutation resulting from a deletion, and two low-priority variants (1 synonymous variant, 1 premature stop codon). The frameshift was confirmed in a separate group of Holsteins from the 1000 Bull Genomes Project that shared no animals with the discovery set. Intraflagellar protein 80 (*IFT80*)-null embryos were generated by truncating the *IFT80* transcript at exon 2, using two guide-RNAs annealed to Cas9 mRNA. Abattoir-derived oocytes were fertilized *in vitro* with a proven high-fertility sire. Embryos were injected at the one-cell stage either with CRISPR-Cas9 complex (n=100) or Cas9 mRNA (control, n=100) before return to culture, and replicated 3 times. *IFT80* is activated at the 8-cell stage, and *IFT80*-null embryos arrested at the 8-cell stage of development, which is consistent with data from mouse hypomorphs and HH2 carrier-to-carrier matings. A

frameshift in *IFT80* on chromosome 1 at 107,172,615 bp (p.Leu381fs) disrupts *wnt* and *hedgehog* signaling, and is responsible for the death of homozygous embryos.

PO0047: Methods: Functional Analysis

NCBI RefSeq Resources for Plant Genomics

Anjana Raina Vatsan, Francoise Thibaud-Nissen and Terence D. Murphy, NCBI/NLM/NIH, Bethesda, MD

National Center for Biotechnology Information (NCBI) generates genome annotation data for a wide variety of plants. It is a non-redundant dataset consisting of coding and non-coding transcripts, associated with a Gene record. There are currently 107 plant species annotated using the NCBI's Eukaryotic Genome Annotation Pipeline. Few of these annotated genomes are in scope for manual curation. Manual curation ensures accurate and full-length representation of nucleotide and protein sequences and helps resolve data conflicts and ambiguities. Gene and protein names are assigned, and publications added, when available. It provides for a more accurate and enriched data set.

Data is available through our Gene resource, RefSeq database and a powerful genome browser called the Genome Data Viewer (GDV). It is also accessible via BLAST and Entrez. The full list of plants that NCBI has annotated is available at: https://www.ncbi.nlm.nih.gov/genome/annotation_euk/all/.

Data can be downloaded from our FTP directory at <ftp://ftp.ncbi.nlm.nih.gov/genomes/refseq/plant/> or accessed from the RefSeq home page at <https://www.ncbi.nlm.nih.gov/refseq/>.

PE0048: Methods: Functional Analysis

Introduction and Overview of Kbase and Joint Genome Institute Plant Program

Robert W. Cottingham, Oak Ridge National Laboratory, Oak Ridge, TN and Jeremy Schmutz, HudsonAlpha Institute for Biotechnology, Huntsville, AL

The Department of Energy Biological and Environmental Research (DOE-BER, www.energy.gov/science/ber) program supports research and facilities to achieve a predictive understanding of complex biological, earth, and environmental systems with the aim of advancing the nation's energy and infrastructure security. The program seeks to discover the underlying biology of plants and microbes as they respond to and modify their environments. This knowledge enables the reengineering of microbes and plants for energy and other applications. BER research also advances understanding of the dynamic processes needed to model the Earth system, including atmospheric, land masses, ocean, sea ice, and subsurface processes.

BER funds both a large scale user facility for plant genomics at the DOE Joint Genome Institute (www.JGI.doe.gov), and an open and collaborative computational resource for predictive systems biology of microbes, plants and their communities called the DOE Systems Biology Knowledgebase (www.KBase.us). Both endeavor to help scientists conduct experiments and analyses in areas such as improving biofuel development, understanding plant model systems, advancing plant comparative science and investigating global carbon cycling. In this annual workshop, speakers present current and ongoing developments in their research enabled by JGI and KBase toward Increasingly large-scale and integrative biology relevant to DOE-BER mission. We will also give a brief introduction describing how to apply for access to the JGI Community Science Program, and how you can use KBase to accelerate your plant genomics research.

JGI enables scientific advances accomplished in collaborative projects through the Community Sequencing Program and the DOE BioEnergy Research Centers. The JGI Plant Program is dedicated to applying advances in genomic technologies for understanding fundamental plant biology through comparative genomics and targeted experiments. Our major goal, in collaboration with plant scientists, is to apply this understanding from genomics to accelerate the improvement and domestication of biofuel crops. The JGI Plant program has produced many of the high-quality reference plant genomes available today and we continue to curate and make available comparative data and analysis via phytozome.jgi.doe.gov. The JGI Plant Flagship genomes are continually improved for accuracy and completeness of the genome sequence and the reference annotation.

KBase hosts a suite of resources, Apps, and workflows designed specifically to explore plant genomics data and model plant-microbe interactions. KBase Apps are available to run quality control on reads, annotate plant transcripts, group orthologous proteins from multiple genomes, analyze differential expression, build gene trees, and more. Data is available from both JGI's Phytozome and MycoCosm that can easily be used for analyses within a KBase Narrative to build metabolic models of your organism or with data from PlantSEED. Create your own workflows by combining Apps or follow tutorials to identify gene families, model plant- microbe metabolic interactions, and analyze RNA-seq data. To learn the RNA-seq Pipeline follow the Narrative Tutorial <http://kbase.us/expression-analysis/>. KBase provides reference guided RNA-seq pipelines for microbial, fungal, and plant genomes obtained from the Illumina platform. Visit kbase.us to learn how KBase can enhance your plant science research.

PO0049: Methods: Functional Analysis

RefEx, a Reference Gene Expression Dataset as a Web Tool for the Functional Analysis of Genes

Hiromasa Ono¹, Shuya Ikeda² and Hidemasa Bono¹, (1)Database Center for Life Science, Mishima, Japan, (2)Database Center for Life Science, Kashiwa, Japan

RefEx (Reference Expression dataset; <http://refex.dbcls.jp>) is a web tool for browsing reference gene expression, which provides access to curated data from several other public databases, with expression levels in forty tissues measured by four well-established gene-expression quantification technologies. "RefEx" contains the data from three kinds of organisms, human, mouse, and rat, obtained from normal tissues and cell lines (556 tissues/cell lines in total) measured by four different methods (EST, GeneChip, CAGE, RNA-seq). All data are acquired from public database including those from FANTOM5 project.

Along with an extensive collection of gene expression data above "RefEx" enables the comparison of the gene expression status in each tissue/cell with the difference among the measurement methods. You can search the data simply by gene name, or gene ontology and family name to obtain the data for certain group of genes. Furthermore, you can select "tissue/cell-specific genes", namely marker genes representing characteristics of the tissue/cell calculated by applying a uniform method to all the accumulated public data by clicking tissue icons in "RefEx" top page.

The search/select result shows the comparison of the relative expression levels among tissues and among the four measurement methods, and the relative expression amount is reflected as a heat-map in the 3D model of the human body. You can also compare annotation information (Gene Ontology, etc.) on functions assigned to genes in the search results. These functions support new knowledge discoveries and hypothesis buildings.

Using "RefEx", researchers can confirm the expression level of the genes of interest in many tissues or cells under normal condition, without bench-top experiments. It is also useful as a tool to know the relationship of genes found in functional analysis to elucidate biological phenomena and interpret research results leading to the development of medicines etc. Thus "RefEx" is expected to contribute a wide variety of life science and medical research as a powerful web tool for gene expression analysis.

The RefEx paper published in Scientific Data. <https://doi.org/10.1038/sdata.2017.105>

PO0051: Methods: Functional Analysis

AnnoScore - Comprehensive Protein Function Prediction Pipeline

Christoph Malisi, Ted Gibbons, Joerg Hagmann and Sebastian J. Schultheiss, Computomics GmbH, Tuebingen, Germany

AnnoScore is an automated computational pipeline for the functional annotation of protein sequences. It was developed to be able to handle large eukaryotic proteomes. It is modular and can integrate results from different functional prediction tools into one easy to use database. In our default setup, AnnoScore uses InterProScan to identify matches in InterPro and its member databases like Pfam, SUPERFAMILY, etc., and transfers the associated annotations from Gene Ontology, KEGG, and Reactome. AnnoScore also identifies the closest homologs in the NCBI

non-redundant (NR) protein database. This is done by an in-house developed reciprocal best hits approach. The proteins are mapped to NCBI NR using the ultra-fast blast-like aligner diamond. For each query, the matched proteins with an alignment bitscore close to the highest score are then mapped back to the input proteome. If the original query protein is among the reverse matches with a score close to the highest, the match is retained. This reciprocal near-best hits approach strikes a balance between using too many and potentially misleading matches as in a forward-only search, and removing informative matches as in a strict reciprocal best hits approach. AnnoScore automatically transfers annotations to the matches from Gene Ontology, NCBI Gene, KEGG, and Reactome.

PE0052: Methods: Functional Analysis

DIY Phenology: Open Source and Affordable Technology Development in Plant Phenotyping

Takanari Tanabata and Sachiko Isobe, Kazusa DNA Research Institute, Kisarazu, Chiba, Japan

Various types of technologies are required in plant phenotyping and therefore the best way of technology development in plant phenotyping is to create optimized technology one by one for each research theme. Plant phenotyping technologies should be kept renewing together with advancing biological research activities. However, it requires lots of developmental resources, hence efficiency gain in technology development is considered as one of the most important factors.

To realize an open-source and affordable technology approach in plant phenotyping, we propose the notion of DIY Phenology by open-source and affordable technology. The idea is to provide information so that users can optimize the tools by themselves and thereby reduce the developmental tasks, as with open-source software.

It is considered that the open-source concept is an essential key of the DIY Phenology, and a development cycle with three actions—develop, share and rebuild—will accelerate the development of digital plant phenotyping. The cycle also reduces the cost of technology development and allows users to establish optimized phenotyping tools based on their own specific demands. It is expected that phenotyping technologies developed based on the concept of DIY Phenology will contribute to the future development of plant science, breeding and agriculture.

This work is supported by JST CREST Grant Number JPMJCR16O1, Japan.

PO0053: Methods: Functional Analysis

"Sal da Terra" Database - Towards a Multiomics Approach to Better Understand Plant Response to Salinity Stress

Manoel Teixeira Souza Junior¹, Vivianny Nayse Belo Silva², Letícia Vieira Rios², Jorge Candido Rodrigues-Neto³, Thalita Massaro Malheiros Ferreira², André Pereira Leão¹, Jose Antonio de Aquino Ribeiro¹, Carlos Antonio Ferreira de Sousa⁴ and Patricia Verardi Abdelnur¹, (1)Embrapa Agroenergy, Brasília - Distrito Federal, Brazil, (2)Graduate Program on Plant Biotechnology - Federal University of Lavras, Lavras - Minas Gerais, Brazil, (3)Institute of Chemistry, Goiânia, Brazil, (4)Embrapa, Teresina - Piauí, Brazil

Salinity is one of the main abiotic stresses affecting agriculture worldwide, and it is referred to as the concentration of inorganic salts that are dissolved in soil solution and/or water. The main examples of cations/anions found in the soil solution are Na⁺, Ca²⁺, Mg²⁺, K⁺, Cl⁻, SO₄²⁻, HCO₃⁻, CO₃²⁻, and NO₃⁻. To estimate salinity, measurements of electrical conductivity, total dissolved solids, osmotic potential, and electrical conductivity of solution saturation extract (ECe) are generally employed. Based on soil chemical properties such as ECe, exchangeable sodium percentage (ESP) and saturated soil paste pH (pHs), the soil is considered saline when ECe>4dS/m, ESP<15% and pHs<8.5. Salinization can result from natural causes (natural or primary salinity) as well as from anthropic actions (secondary or man-induced salinity). Primary salinization is induced by weathering and deposition of oceanic salts; while secondary salinization occurs commonly due to deforestation and irrigation. FAO (2019) has estimated that out of the 1.5 billion hectares of rain-fed and the 230 million hectares of irrigated agricultural land in the World, 2.1% and 19.5% are affected by salt, respectively. Plants are classified into two groups accordingly to their response to salinity stress: glycophytes and halophytes. Glycophytes are unable to complete their life cycle in high salinity environments, while halophytes can survive and reproduce in saline environments with a concentration of

NaCl \geq 200mM. The advances in high-throughput omics technologies in the last 10 years or so, and the significant reduction in their cost, gave rise to many initiatives of multi-omics approaches to better understand traits such as response to abiotic and/or biotic stresses, generating an unparalleled amount of data. Our group has embraced such challenge with the aim of better understand the differences among the responses of a glycophyte plant (oil palm – *Elaeis guineensis* Jacq.) and two supposedly halophyte plants [purslane (*Portulaca oleracea* L.) and gliricidia (*Gliricidia sepium* (Jacq.) Steud.); and develop technologies to mitigate the effects of salinity in economically important glycophyte plants. We started by setting up and validating the bioassays used to evaluate the morphophysiological responses of these plants to salinity stress. Once validated, these bioassays were used to collect phenomics data (daily evapotranspiration rates, gas exchange measurements, leaf temperature, chlorophyll content index, chlorophyll fluorescence measures, etc.). Plant material (roots and leaves) were also collected for transcriptomics (RNASeq - mRNAs and regulatory ncRNAs, separately), metabolomics (UHPLC-ESI-MS), and proteomics data. Additional data are being gathered, such as physicochemical properties of the substrate and mineral composition of roots and leaves of plants, biomass accumulation, and light and scanning electron microscopy pictures of leaves and roots. Purslane and gliricidia have shown tolerance to salt stress, manifested through different symptoms. Gliricidia has shown a highly interesting recovery phenotype, which is now under further characterization. A bioassay was also developed to evaluate the morphophysiological responses of *Setaria viridis* L., a model plant for C4 species, to salinity stress. The database built as a result of this multi-omics approach was named “Sal da Terra”.

PE0054: Methods: Functional Analysis

Strategies for Improving Plant Drought Tolerance for a Hotter, Drier World

John C. Cushman, University of Nevada - Reno, Reno, NV

The climate crisis driven by increasing greenhouse gas emissions threatens to increase the frequency, duration, and geographical distribution of droughts over global landmasses during the 21st century. As crops become more sensitive to heat and drought stress, the rates of grain yield increases are projected to decline with losses worsening in the latter part of this century. In order to develop novel strategies for improving, we have investigated a specialized form of photosynthesis known as crassulacean acid metabolism (CAM) that is present in more than 6% of vascular plant species along with a suite of co-adapted traits (e.g., tissue succulence, water capture and storage strategies, thick cuticles, enhanced epicuticular wax accumulation, reduced stomatal density, increased stomatal responsiveness, and rectifier-like roots) that might serve to improve the adaptability of plants to hotter and drier climates. CAM increases water-use efficiency (WUE) and reduces water demand through the use inverted stomatal behavior coupled with a temporal CO₂ pump with nocturnal CO₂ uptake and concentration. Thus, introducing the CAM pathway into C₃ photosynthesis plants (CAM Biodesign) is expected to confer enhanced photosynthetic performance and WUE. Current steps achieved to date for CAM Biodesign will be summarized including subcellular localization and phenotypic analysis of overexpressing 14 individual ice plant C₄-cycle genes, mesophyll-specific, circadian clock-controlled promoter mining, vector set construction for multi-gene circuit assembly, and the phenotypic effects of engineering a four-component carboxylation module in *Arabidopsis*. In addition to engineered CAM, we have explored the effects of increasing tissue succulence on plant growth, productivity, drought acclimation, and salinity stress tolerance in *Arabidopsis*. Increasing cell size resulted in a 2–3-fold increase in leaf succulence (defined as the water content of the leaf/leaf area) with a corresponding decrease in stomatal density and aperture, which resulted in a 1.5–2.8-fold increase in instantaneous WUE and a 2.1–2.3-fold increase in integrated WUE compared to control lines. This improved WUE resulted in significant improvements in aerial biomass and seed yield under both acute and chronic water-deficit stress. Enhanced tissue succulence also resulted in significant increases in aerial biomass and seed yield under both acute and chronic salinity stresses due to a reduction in the effective Na⁺ and Cl⁻ concentrations within leaves and reduced Na⁺ uptake. These results indicate that new approaches to improving drought attenuation and salinity tolerance are possible through relatively small changes in leaf anatomy and architecture.

PO0055: Methods: Functional Analysis

Dynamic Quantification and Annotation of Regulatory Elements and Gene Bodies Using Total RNA

Sascha Duttke, Max W. Chang, Sven Heinz and Christopher Benner, University of California, San Diego, La Jolla, CA

Capturing active transcription initiation is critical for studying gene expression and functional annotation of regulatory elements and genes in genomes. To facilitate such analysis in fresh or frozen tissues and complex organisms where nuclei isolation or genetic manipulations are not readily feasible, we developed capped-small RNA-seq (csRNA-seq), which uses total RNA as starting material to detect transcription start sites of both stable and unstable RNAs at single-nucleotide resolution. csRNA-seq is highly sensitive to acute changes in transcription and identifies an order of magnitude more regulated transcripts than does RNA-seq. Interrogating tissues from species across the eukaryotic kingdoms identified unstable transcripts resembling enhancer RNAs, pri-miRNAs, antisense transcripts, and promoter upstream transcripts in multicellular animals, plants, and fungi spanning 1.6 billion years of evolution. Integration of total RNA-seq data further facilitates approximating transcript stability as well as accurate annotation of genes and transcribed regulatory elements such as enhancers. Our findings show that total RNA is sufficient to identify and annotate genes as well as transcribed regulatory elements and capture the dynamics of initiated stable and unstable transcripts at single-nucleotide resolution in eukaryotes.

PE0056: Methods: Functional Analysis

Characterizing Genetic Variations for Wheat Streak Mosaic Virus (WSMV) Resistance in Winter Wheat

Yucong Xie¹, Karl Ravet² and Stephen Pearce², (1)Colorado State University, Fort Collins, CO, (2)Colorado State University

Wheat streak mosaic virus (WSMV) (Genus *Tritimovirus*; Family *Potyviridae*), transmitted by wheat curl mite (WCM), causes serious yield loss to wheat and other cereals in the Great Plains region of the United States. Host plants with genetic resistance to WSMV are an effective way to control this disease, but only limited options are available to breeders. The goal of this project is to identify novel and durable genetic variants and candidate gene(s) associated with WSMV resistance.

A field-based GWAS with 597 advanced wheat lines identified a QTL for WSMV resistance on chromosome arm 3BS that corresponds to the previously identified *Wsm2* locus (Dhakal et al. 2018). We validated this QTL in a doubled haploid population (n=136, LOD = 15.08, $P < 0.0001$) derived from parental lines Snowmass (*Wsm2*+, WSMV resistant) and Antero (*Wsm2*-, WSMV susceptible) in growth chamber experiments with manual WSMV inoculation. The flanking markers for this QTL were mapped to a 1.58 Mb region based on the RefSeq v1.0 genome assembly, which contains 68 genes including 38 high-confidence and 30 low-confidence gene models.

To identify the genetic variants underlying *Wsm2* we mapped exome sequence data from the phenotypically different parents with or without *Wsm2* region to identify polymorphisms. Additionally, we utilized wheat pan-genome and comparative genomic approach for identification of novel variants from whole genomic level and make inference for the genome information of the WSMV resistant and susceptible parents that have not been sequenced yet. We also conducted a RNAseq experiment to identify differentially expressed genes (DEGs) between WSMV resistant and susceptible wheat varieties following WSMV infection. The combination of genomic and transcriptomic approaches facilitates the characterization of novel genetic variants responsible for WSMV resistance.

PO0057: Methods: High-throughput Methods

High Throughput Agrigenomic SNP Genotyping Using the Applied Biosystems™ Axiom™ Genotyping Solution

Wilson S. Lew, Claudia Diaz, John Mundaden, Devin Nguyen, Nobuo Ogawa, Jacob Peete, Yumi Suh, Antonita Yan, Chi-Sou Yu and Michael Shapero, Thermo Fisher Scientific, Santa Clara, CA

High density DNA microarrays play an important role in agrigenomic research as climate change, population growth, and urbanization threaten the ability of farmers to meet the world's food demands. Microarrays enable accurate, cost-effective genotyping of variants that include single nucleotide, insertion/deletion and multiallelic

polymorphisms. Applications include genomic selection, marker-assisted selection (MAS) and marker-assisted breeding (MAB), parentage and characterization of genetically modified organisms (GMOs). Here an overview of recent advances related to scalable, high throughput workflows using the Axiom solution is presented.

The Axiom solution currently enables complete automation of DNA target preparation on several liquid handling instruments. A new high throughput workflow that scales to 20,000 samples per week with a focus on cost-effective lab configurations has been developed. This workflow emphasizes smaller, simpler devices coupled with bulk reagent packaging. Customizable content can be used for genotyping of any species, genome size, or ploidy level in a 96-array layout (up to 650,000 variants) or 384-array layout (up to 50,000 variants). Additionally, a protocol enabling 48 hour turnaround time from sample to answer is described.

In summary, new workflows to the Axiom solution will drive expansion of microarrays into research on complex genetic traits in plants or animals. Scalable lab automation extends the platform capabilities to genotype SNPs in a single assay with a sample throughput consistent with the needs of both breeders and farmers who are employing new genomic strategies in order to use fewer environmental resources, antibiotics and pesticides to develop higher-producing livestock, poultry, and crops.

PE0058: Methods: High-throughput Methods

Efficient Genotyping and Powerful Data Analysis with the Agilent ZAG DNA Analyzer System

Whitney A. Pike¹, Solange Borg², Chava Pocernich¹ and Kyle David Luttgarm¹, (1)Agilent Technologies, Ankeny, IA, (2)Agilent Technologies, San Diego, CA

Genotyping by gel electrophoresis can be a long and cumbersome task. Here, we present an alternative method for streamlining high-throughput fragment analyses. A segregating line of *Arabidopsis thaliana* LOH1 knockout seeds was grown and harvested at three weeks for genomic DNA extraction. The DNA was amplified with a three primer PCR method to screen for mutations. To determine the genotype, the PCR products were analyzed using the ZAG DNA Analyzer system with the ZAG 110 dsDNA kit (35-5000 bp). With this system, 96 samples were separated simultaneously using parallel capillary electrophoresis, and the results analyzed using ProSize data analysis software. The resulting digital gel image provided easy visualization of the results, with the wildtype sample displaying a large band at approximately 1,200 bp, the homozygous mutants a smaller band at approximately 500 bp, and the heterozygous plants showing both bands. Advanced options provide unbiased analysis of sizing and can flag samples based on user-defined criteria. Using these options, we created flags for the presence and/or absence of a band at each of the size ranges (1100 +/- 150 and 500 +/- 50). The results were exported in a table indicating the genotype of each sample, allowing for sample calls to be made without human error. The exported results matched the calls that were made by interpretation of the gel images. This high-throughput method of genotyping analysis with the ZAG DNA Analyzer system can easily be applied to other model organisms, including maize, mice, and zebrafish.

PO0059: Methods: High-throughput Methods

Long DNA Technologies Evaluation For Structural Variations Detection: Nanopore ONT Sequencing vs BioNano Genomics Optical Mapping

Aur lie Canaguier¹, Erwan Denis², Romane Guilbaud¹, Ghislaine Magdelenat³, Caroline Belser³, Benjamin Istace³, Corinne Cruaud³, Arnaud Lemainque³, Patrick Wincker³, **Marie-Christine Le Paslier¹**, Patricia Faivre Rampant¹ and Valerie Barbe², (1)Etude du Polymorphisme des G nomes V g taux (EPGV), INRA, Universit  Paris-Saclay, 91000 Evry, France, (2)G nomique M tabolique, Genoscope, Institut de biologie Fran ois Jacob, Commissariat   l'Energie Atomique (CEA), CNRS, Universit  Evry, Universit  Paris-Saclay, Evry, France, (3)Genoscope, Institut de biologie Fran ois Jacob, Commissariat   l'Energie Atomique (CEA), Universit  Paris-Saclay, Evry, France

Structural Variations (SVs) regroup very diverse genomic rearrangements (with a length often more than 50bp). Their identification remains in infancy in plants mainly due to the limited useful genomic resources. Today, last technologies based on long DNA fragments offer new possibilities to identify such polymorphisms.

In our study, long-read sequencing obtained with the MinION instrument (Oxford Nanopore Technologies (ONT)) and optical maps produced by the Saphyr device (BioNano Genomics), were evaluated on their ability to identify SVs between two *Arabidopsis thaliana* ecotypes, Columbia-0 (Col-0) and Landsberg erecta-1 (Ler-1).

Corrected and trimmed Col-0 and Ler-1 ONT sequences were assembled with SMARTdenovo. MUMmer tools were used to align the assemblies on public references of each ecotype and for the SVs detection. In parallel, we used the BioNano SV detection tool which identifies the SVs by comparing the maps obtained with the DLE-1 enzyme with the *in silico* digested reference genome maps. The ONT and BioNano SVs were considered as the same if their absolute positions overlapped.

For all the comparisons performed in our study, more than 90% of BioNano SVs were detected with ONT and ~70% of ONT SVs were obtained with BioNano. However, The SV > 1000b were identified by both technologies. ONT SVs were smaller and less reliable in complex region whereas BioNano SVs were longer with more confidence. Challenges on wet lab part, in particular extraction of high molecular weight DNA, are still relevant.

Keywords : Structural Variations, Long Read Sequencing, Oxford Nanopore Technologies, Optical Mapping, BioNano Genomics, *Arabidopsis thaliana*

PE0060: Methods: High-throughput Methods

High-Throughput Method of Shearing DNA using Bead Mill Technology for ChIP

Brandon Easparro¹, Gabriella Ryan^{1,2}, Taylor White^{1,2}, Caleb Proctor¹ and Rodney Nash^{1,2}, (1)Omni International, Kennesaw, GA, (2)Georgia State University, Atlanta, GA

Chromatin Immunoprecipitation (ChIP) is key to understanding DNA-protein interactions. This includes the mapping transcriptional factors and histone modifiers. Recent advancements have shown a transition from ChIP-chip experiments to ChIP-seq experiments, partly due to a reduction in materials cost, an increase in genomic coverage, as well as a reduction in required starting material. However, preparation of the chromatin is still prepared by either shearing cross-linked chromatin by sonication or native chromatin micrococcal nuclease digestion. These options require longer incubation times and lower throughput, resulting in overall longer preparation. Bead milling provides an alternative of shearing DNA through the use of high impact beads. Through the use of specific beads, DNA shearing is achievable. Here we report the comparison of bead milling to ultrasonic shearing of DNA for ChIP analysis.

PO0061: Methods: High-throughput Methods

Versatile, Robust, Low Cost Genomic DNA Extraction Solution for Use across Multiple Sample Types and Downstream Genomic Platforms

Quoc Hoang¹, Lillian Manley¹, Calvin Gunter², **Srinivas Chadaram¹**, Mohini A. Patil³, Tao Lei³ and Rick Conrad¹, (1)Thermo Fisher Scientific, Austin, TX, (2)Thermo Fisher Scientific - Animal Health, Austin, TX, (3)Thermo Fisher Scientific, Santa Clara, CA

Obtaining a low cost and robust method for high throughput sample preparation upstream of genomic platforms is important for laboratories that need to run hundreds or thousands of samples a day. Many laboratories need to run a diverse set of sample types such as blood, hair, semen, and tissue. Low cost workflows often involve a crude lysate that can lead to poor results. The MagMAX™ CORE AgGenomic DNA Extraction Kit was designed to be a robust genetic DNA extraction kit that works with a diverse range of sample types yielding DNA that is suitable across multiple genomic platforms. This kit uses magnetic beads in conjunction with the KingFisher™ Flex Purification System to extract DNA. The total processing time from sample to purified DNA is around 1 hour. Here, we examine DNA isolated from bovine blood, blood cards, raw and extended semen, ear notch, and hair follicles isolated using the MagMAX™ CORE AgGenomic DNA Extraction Kit compared to a more expensive on market magnetic bead-based isolation kit. For down stream applications we tested capillary electrophoresis on an ABI 3500 Genetic Analyzer, Applied Biosystems™ Axiom™ Genotyping Arrays, and targeted GBS with AgriSeq™ HTS Library kits on an Ion GeneStudio™ S5. The data shows the MagMAX™ CORE AgGenomic DNA Extraction Kit was able to extract DNA from all samples types tested and is compatible with all the genetic platforms tested.

PE0062: Methods: High-throughput Methods

Allegro Targeted Genotyping V2 on the NGS DreamPrep™ Automation Platform

Michael Benway, James Anderson, Stephanie C. Bruns and Joe Don Heath, Tecan Genomics, Inc., Redwood City, CA

High throughput sequence-based genotyping is an indispensable tool in crop and livestock breeding. As the practice continues to expand to more organisms and higher sample numbers, there is growing pressure on increasing throughput and decreasing costs. Here we describe and present results from the complete end-to-end automation of the most recent release of the Allegro Targeted Genotyping V2 by Tecan Genomics on the NGS DreamPrep™ liquid handling automation platform. The Allegro V2 system features workflow improvements that increase throughput and reduce costs while maintaining the flexibility of design and quality of data of the innovative NGS-based targeted genotyping system. Using the patented Single Primer Enrichment Technology (SPET) for DNA approach to specifically target SNPs of interest, Allegro Targeted Genotyping provides information-rich sequencing data, by capturing a SNP-specific data point for every on-target sequencing read. On the NGS DreamPrep system, Tecan's Fluent® is able to process up to four plates of 96 DNA samples to pooled targeted genotyping NGS libraries ready for sequencing in less than 24 hours, thus greatly increasing sample throughput and decreasing both turnaround time and hands-on time. Results from both plant and human samples will be presented demonstrating high on-target rates, high SNP call rates and excellent concordance with SNP frequencies from orthogonal genotyping approaches such as microarrays and qPCR.

PO0063: Methods: High-throughput Methods

A Quick and Easy DNA Preparation from Crops with Lab-Made Magnetic Nanoparticles

Haichuan Wang¹, Xueqi Zhao², Li Tan¹, Ravi Murali¹ and David L. Hyten³, (1)UNL, Lincoln, NE, (2)3MOE Key Laboratory for Nonequilibrium Synthesis and Modulation of Condensed Matter, School of Science, Xi'an Jiaotong University,, Xi'an, China, (3)University of Nebraska-Lincoln, Lincoln, NE

A quick and inexpensive high-throughput method for isolating genomic DNA from three cereal crops, maize (*Zea mays*), soybean (*Glycine max*) and wheat (*Triticum aestivum*) leaf tissues, and soybean and wheat seeds is reported. This method could use either collected in 96-well plate (s) or 1.7 ml centrifuge tube(s) for the genomic DNA isolation. The DNA isolation step is performed with magnetic nanoparticles less than 2 hrs using traditional CTAB (hexadecyltrimethylammonium bromide) buffer, which enables purification from multiple samples in a single run with less steps. Furthermore, the yields of genomic DNA range from 2.92.55-1989 ng per sample and 7.2-60.3 ng per seed (soybean) and per wheat seed(s) varying with the number of leaf discs or seed(s) applied, respectively. The DNAs prepared using this method have been further tested and proved to be suitable for downstream analysis, such as polymerase chain reaction (PCR) and molecular inversion probe (MIP) for the identification of alleles in diverse genetic and breeding approaches, such as marker-assisted selection and genetic fine mapping. The established method is fast, simple, reliable, and environmentally friendly.

PE0064: Methods: High-throughput Methods

Plant Regulomics: A Data-Driven Interface for Retrieving Upstream Regulators from Plant Multi-Omics Data.

Yi-Jing Zhang, Institute of Plant Physiology and Ecology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai, China

High-throughput technology has become a powerful approach for routine plant research. Interpreting the biological significance of high-throughput data has largely focused on the functional characterization of a large gene list or genomic loci that involves the following two aspects: the functions of the genes or loci and how they are regulated as a whole, i.e. searching for the upstream regulators. Traditional platforms for functional annotation largely help resolving the first issue. Addressing the second issue is essential for a global understanding of the regulatory mechanism, but is more challenging, and requires additional high-throughput experimental evidence and a unified statistical framework for data-mining. The rapid accumulation of 'omics data provides a large amount of experimental data.

We here present Plant Regulomics, an interface that integrates 19 925 transcriptomic and epigenomic data sets and diverse sources of functional evidence (58 112 terms and 695 414 protein–protein interactions) from six plant species along with the orthologous genes from 56 whole-genome sequenced plant species. All pair-wise transcriptomic comparisons with biological significance within the same study were performed, and all epigenomic data were processed to genomic loci targeted by various factors. These data were well organized to gene modules and loci lists, which were further implemented into the same statistical framework. For any input gene list or genomic loci, Plant Regulomics retrieves the upstream factors, treatments, and experimental/environmental conditions regulating the input from the integrated 'omics data. Additionally, multiple tools and an interactive visualization are available through a user-friendly web interface.

Plant Regulomics is available at <http://bioinfo.sibs.ac.cn/plant-regulomics>.

PO0065: Methods: High-throughput Methods

OneKK: A High Throughput Seed Phenotyping Android Application

Trevor W. Rife¹, Megan Calvert², Chaney Courtney², Mitchell Neilsen² and Jesse Poland¹,
(1)Department of Plant Pathology, Kansas State University, Manhattan, KS, (2)Kansas State University, Manhattan, KS

Seed size and morphology has an important effect on the end uses of crops. Rapidly measuring morphological phenotypes and utilizing this information for indirect selection within breeding programs could lead to increased yields and improved end use quality. High-throughput approaches are useful for many crops since they can provide rapid and accurate measurements, but commercial solutions are expensive and outside the budgets of most plant breeding programs. OneKK, a new app that runs on Android smartphones and tablets, makes rapid seed phenotyping accessible, portable, and cost effective. OneKK uses an established algorithm to calculate length and width and a novel watershed algorithmic approach to estimate the number of seeds within the image – even when seeds are immediately adjacent. To validate the accuracy of OneKK, seeds from common crops were manually measured for length and width. The same samples were processed using OneKK to measure the average length, average width, and sample count. A high correlation between both morphological measurements and seed counts was observed, and measurements from OneKK were collected considerably faster. To validate the utility of OneKK for genomic research, the Synthetic/Oyata doubled haploid wheat population was utilized for QTL mapping. Seed measurements taken with OneKK were successfully used to map a QTL for seed length and width. OneKK is a free and flexible app that will provide all plant breeding and genetics research programs with the data necessary to perform both phenotypic selection and genomic analysis.

PE0066: Methods: High-throughput Methods

araDeepopsis: A Transfer Learning Approach to High-Throughput Plant Phenotyping

Patrick Huether, Niklas Schandry and Claude Becker, Gregor Mendel Institute of Molecular Plant Biology, Vienna, Austria

In the post-genomic era, phenotypic data has become the limiting factor in genomic studies, as genotype data are readily available.

The advent of new sequencing technologies has led to a plethora of plant genome resources available to the scientific community. While panels such as 1001-Genomes project in *Arabidopsis thaliana* or the 3000-Rice-Genomes project provide extensive genotype information, a severe bottleneck currently is to collect high-throughput phenotype measurements. A major challenge lies in maximizing the accuracy of such measurements, while at the same time retaining as many data points as possible, in order to faithfully relate phenotype and genotype.

Existing methods for (semi-)automated plant phenotyping are often image-based and have limitations when it comes to throughput, scalability, or robustness towards changes in conditions during image acquisition. As these methods often rely on color channel information, they are very sensitive to changes in illumination, plant color, or background configuration.

Here, we present *araDeepopsis*, an open-source tool that allows robust and versatile measurement of phenotypic traits from image data using a transfer learning approach. *araDeepopsis* is built upon the publicly available convolutional neural network DeepLabv3+ and is able to efficiently assess plant phenotypes.

We show how such a model, in combination with scikit-image, can extract biologically relevant phenotypic traits from images, independent of phenotyping platform, background, plant health and developmental stage. For *A. thaliana* rosettes, *araDeepopsis* reached a segmentation accuracy of 97% after training on only 300 manually annotated images. Because of its open-source character and robust training on small training sets, users can expand *araDeepopsis* to their preferred plant species or phenotype with little time investment or prior knowledge.

PO0067: Methods: High-throughput Methods

A High-Density Environmental Sensing System Revealed Effect of Point-Specific Environmental Differences on Phenotypes of Greenhouse Tomato

Eiji Yamamoto^{1,2}, Takanari Tanabata³, Kenta Shirasawa¹ and Sachiko Isobe³, (1)Kazusa DNA Research Institute, Kisarazu, Japan, (2)JST PRESTO, Kawaguchi, Japan, (3)Kazusa DNA Research Institute, Kisarazu, Chiba, Japan

Plant phenotypes are highly sensitive to environmental changes. This disturbs not only agricultural production but also phenotype evaluation for genetic and physiological studies. Capturing environmental information on each plant in an experimental field has a potential not only for analyzing environmental effect on plant growth but also for accurate evaluation of genetic potential. We have developed a high-density environmental sensing system that records solar radiation, air temperature, humidity, soil temperature and soil moisture on each plant for every 10 minutes. The system was installed in a big-fruited tomato greenhouse with hydroponic cultivation system. Observation of environmental changes during a growing period revealed point-specific environmental differences, especially in soil moisture. The soil moisture difference among the plants was associated with incidence of blossom-end rot, a major physiological disorder in tomato. In this study, we show an example of how to use the environmental data for analyses of the effect on plant growth and the genetic evaluation.

PE0068: Methods: High-throughput Methods

Evaluation of Tractor-Based High Throughput Phenotyping Methods for Use in Tomato and Pepper Breeding

Bryce A. Kubond¹, Vivian Vuong¹, Amanjot Kaur¹, Paul Bosland², David C. Slaughter¹ and Dina A. St. Clair¹, (1)University of California, Davis, Davis, CA, (2)New Mexico State University, Las Cruces, NM

High Throughput Phenotyping (HTP) technologies are being developed to address the phenotyping bottleneck in crop breeding. In vegetable crops, including tomato and pepper, there has been limited development of HTP for in-field phenotyping. At UCD, a tractor-based HTP system that includes visual spectrum cameras, infrared cameras, time of flight sensors, and RTK-GPS was developed and is being evaluated for use in tomato and pepper breeding. Both manual phenotype data and HTP robot phenotype data (horticultural and water stress related) was collected in 2017 and 2018 on plots of pepper and tomato genotypes for use as a training data set. A phenotypically diverse set of 30 tomato genotypes and 16 pepper genotypes were grown for HTP experiments in 2017-2018. In 2019, a new set of 40 tomato genotypes and 24 pepper genotypes were grown for evaluating HTP methods developed in the prior two years. Manual phenotype data for 9 tomato traits and 3 pepper traits were collected in each of the three years. Significant statistical differences among genotypes were identified for each trait in the manually collected phenotype data. Analysis of HTP robot collected data and comparison of manual phenotype data to HTP robot data is in progress.

PO0069: Methods: High-throughput Methods

Development and Validation of a Low-Cost, Rapid NIRS-Based Phenotyping Approach for Improving Cassava Root Quality Traits

Jenna M. Hershberger¹, Bryan Ellerbrock², Lukas A. Mueller², Robert Kawuki³, Eder Oliveira⁴, Ismail Rabbi⁵ and Michael A. Gore¹, (1)Plant Breeding and Genetics Section, School of Integrative Plant Science, Cornell University, Ithaca, NY, (2)Boyce Thompson Institute, Ithaca, NY, (3)National Root Crops Resources Research Institute, Namulonge, Uganda, (4)Embrapa, Brasília, Brazil, (5)International Institute of Tropical Agriculture (IITA), Ibadan, Nigeria

Over 800 million people across the tropics rely on cassava as a major source of calories. While the dry matter content (DMC) of this starchy root is important for both growers and consumers, characterization of DMC by traditional methods is time-consuming and laborious. Near-infrared spectroscopy (NIRS) has been proposed as an alternate phenotyping method, but while it is highly predictive of DMC in cassava roots, spectrometers that have been validated for high accuracies are prohibitively expensive, limiting their usefulness. For this reason, we investigated the use of a low-cost, handheld NIR spectrometer (SCiO, ConsumerPhysics) for field-based DMC prediction in cassava roots. Pilot investigation into the predictive effects of preprocessing techniques, number of roots sampled per plot, and within-root sampling location were used to develop a scanning and sampling protocol. Following this protocol, oven-dried measurements of DMC were paired with scans of roots of diverse clones from IITA (Nigeria), NaCRRI (Uganda), and Embrapa (Brazil) and grouped into training and test sets based on prediction scenarios common to plant breeding programs. Partial least squares regression models were evaluated for predictive ability, which ranged from $R^2=0.51-0.85$ depending on the cross-validation scheme. With appropriate calibration, this spectrometer will allow for field-based collection of spectral data with a smartphone for accurate DMC prediction, a step that could be easily integrated into existing harvesting workflows of cassava breeding programs. These and other NIRS models will be hosted on BreedBase to facilitate further analysis and incorporated into the PhenoApps suite of Android applications for plant phenotyping.

PE0070: Methods: High-throughput Methods

Diversity Study in Citrus With High-Density SNP Array Data

Yoko Eck, Sergio Pietro Ferrante and Mikeal L. Roose, University of California, Riverside, CA

Citrus is one of the most widely cultivated and economically valued fruit tree crops in the world. Given the complex ancestry of Citrus, analysis of germplasm with low-cost, high throughput tools such as genome-wide SNP arrays can significantly reduce the time breeders take to screen and characterize germplasm and to identify the genes responsible for traits of interest. We analyzed genetic diversity in *Citrus* with high-density SNP genotype data generated by a recently developed SNP genotyping array for *Citrus*, Axiom™ Citrus Genotyping Array (Affymetrix, Inc.) (58K autosomal and 500 Chloroplast SNPs). Concordance analysis of 927 named accessions in the UCR citrus variety collection (CVC) was used to identify clonally-derived/near-identical accessions prior to the diversity analyses. Ho, He, PIC (polymorphic information content) and percent polymorphic markers were calculated using array data with 927 unselected accessions, the reduced set of 399 accessions in which clonal groups are represented by one sample each, and 36 accessions included in the variant discovery panel used to design the array. In addition, we performed PCA, admixture, treemix, phylogeny network analyses and graphical genotyping showing ancestry specific loci in selected accessions. We used the most stringent PolyHighResolution (PHR) loci as classified by Axiom™ Analysis Suite in both analyses. Admixture analysis with 399 accessions shows clustering into citron, mandarin, pummelo, trifoliolate and kumquat/microcitrus/papeda at $K=5$ and papeda separates into its own cluster at $K=7$. Our results show high degree of reticulation events in citrus.

PO0071: Methods: High-throughput Methods

Aerial Drone Imagery Database and Prediction of End-of-Season Traits in Maize

Nicolas Morales, Boyce Thompson Institute, Ithaca, NY; Cornell University, Ithaca, NY

High-throughput phenotyping requires a well-structured and easy-to-use data management and processing pipeline. Critical to this effort is integration of ground-truth measurements, such as plant height, yield, and disease severity scores, and existing genotypic resources with the extracted image features. Furthermore, large community efforts, research projects, and breeding programs collecting image data ideally should leverage a standardized and centralized web-database for sharing and combining data. To meet these needs, we present ImageBreed for storing and processing drone images, both RGB and multi-spectral, into the BreedBase family of open-source web databases. BreedBase is used by dozens of plant breeding and genetics projects, including <https://cassavabase.org>

and <https://solgenomics.net>, with plans to scale out to 50 to 100 USDA-ARS breeding programs through the Breeding Insight Platform (BIP). This system allows a researcher to login to the website, upload their field experiment information, upload their ground-truth phenotypic measurements, upload their drone image(s), stitch the aerial drone images into an ortho-mosaic if required, calculate vegetative indices (NDVI, TGI, VARI), remove background soil from the image, define and save plot-level polygon images, calculate and save plot-level zonal-statistics phenotypes, and correlate those extracted phenotypes to the ground-truth phenotypic measurements. Also presented are algorithms for predicting end-of-season traits such as grain yield or grain moisture directly from plot-level aerial images using longitudinal convolutional neural networks (LSTM CNN); these algorithms are trained on imaging events spanning several years and locations. The system is fully operational at <http://imagebreed.org>.

PE0072: Methods: High-throughput Methods

Drone Phenotyping Enabled the Discovery of QTLs Associated with Flower Opening Time in Lettuce

Rongkui Han¹, Andy J.Y. Wong¹, Zhehan Tang¹, Fermin Zurissaday Banuelos-Gonzalez¹ and Richard Micheltore², (1)University of California, Davis, Davis, CA, (2)The Genome Center, University of California, Davis, CA

Flower opening and closure are traits of reproductive importance in all angiosperms, as they determine the success and self- and cross-pollination events. Cultivated and wild lettuce, *Lactuca spp.*, have single-opening flowers, which expand their yellow petals on the ray florets of their composite inflorescences once only, before their prolonged closure that terminates when the seeds reach physiological maturity. Lettuce flowers open and close in a circadian fashion, following a tight diurnal schedule that dictates a specific, highly predictable flower opening time for each individual lettuce genotype. Existing variations in flower opening time have been observed in lettuce in the past. Nevertheless, the transient nature of this phenotype has rendered it a difficult target for genetic studies. In this experiment, an F₆ *L. sativa* x *L. serriola* Recombinant Inbred Line (RIL) population that segregates for flower opening time was used to perform Quantitative Trait Loci (QTL) mapping for this phenotype. The flower opening time trait value was scored using time-course image series obtained by flying a drone mounted with a high-resolution multi-spectral camera over the experimental field in 30-min or 1-hour intervals. Flower pixels were identified from the images using custom RGB value thresholds. We developed a Bayesian statistic algorithm that inferred flower opening time for individual genotypes from time-stamped per-plot flower pixel count data, using a Hamiltonian Markov Chain. The inference results generated by the algorithm correlated strongly with results of manual phenotyping. Three QTLs on Chromosomes 2, 8 and 9 were identified to explain 28% of the phenotypic variation in flower opening time.

PO0073: Methods: High-throughput Methods

Gene Expression Analysis Associated with Salinity Stress in Flax/Linseed (*Linum usitatissimum*)

Manisha Sharma, ICAR-National Institute of Plant Biotechnology, New Delhi, India, Rhitu Rai, National Research Centre on Plant Biotechnology, New Delhi, India, Rohini Sreevathsa, Indian Centre of Agricultural Research, National Research Centre on Plant Biotechnology, New Delhi, India, Raju Datla, Global Institute for Food Security, Saskatoon, SK, Canada and Prasanta K Dash, ICAR-National Research Centre on Plant Biotechnology, New Delhi, India

Salinity is one of the most important environmental factors that limit crop growth, development and productivity. Globally, almost 20% of the world's cultivated land and nearly 50% of irrigated land suffer from saline stress. Saline stress has extensively studied in model plant Arabidopsis. However, in the field crop flax (*Linum usitatissimum* L.) the molecular and physiological mechanisms of salt tolerance are yet to be elucidated. Our study involved comprehensive analysis of flax seedlings to analyze expression perturbation during salinity stress. We carried out physiological and whole genome transcriptomic study of flax roots and shoots exposed to a 6 h and 12 h of salinity treatment. A total of 6134 differentially expressed genes (DEGs) were identified in shoot and root tissues together. Further, in-depth analysis of tissue specific DEGs revealed that gene groups involved in ion transport such as sodium proton antiporter (NHX), osmotic stress, ROS homeostasis and calcium signaling are involved in ionic homeostasis in flax seedlings. Further based on sequence homology search, a key family of Na antiporters in flax-LusNHX was identified and their expression was studied. In conclusion, our results suggest that the differential

expression of stress and transport related genes might be contributing to fine-tune the ionic and ROS homeostasis in flax tissues.

PE0074: Methods: High-throughput Methods

Methylation Content Sensitive Enzyme ddRAD (MCSeEd): A Reference-Free, Whole Genome Profiling System to Address Cytosine/ Adenine Methylation Changes

Emidio Albertini, University of Perugia, Perugia, Italy

Methods for investigating DNA methylation nowadays either require a reference genome and high coverage, or investigate only CG methylation. Moreover, no large-scale analysis can be performed for N⁶-methyladenosine (6mA). Here we describe the methylation content sensitive enzyme double-digest restriction-site-associated DNA (ddRAD) technique (MCSeEd), a reduced-representation, reference-free, cost-effective approach for characterizing whole genome methylation patterns across different methylation contexts (e.g., CG, CHG, CHH, 6mA).

This MCSeEd technique was tested in two maize experimental systems: (i) leaves of a commercial maize hybrid grown under normal irrigation (well watered; WW) and under drought stress (DS), and collected 60 days after sowing (DAS); and (ii) shoots and roots of the inbred line B73 collected at 5 DAS. The relative methylation changes estimated by MCSeEd for differentially methylated positions (DMPs) and differentially methylated regions (DMRs) clearly discriminated between these samples (i.e., WW vs DS; B73 shoots vs roots) with both genome-dependent and genome-independent approaches. The DMRs identified by MCSeEd showed gene enrichments that were related to the experimental system under investigation.

MCSeEd can detect genetic variations among hundreds of samples. MCSeEd is based on parallel restrictions carried out by combinations of methylation insensitive and sensitive endonucleases, followed by next-generation sequencing. Moreover, we present a robust bioinformatic pipeline (available at <https://bitbucket.org/capemaster/mcseed/src/master/>) for differential methylation analysis combined with single nucleotide polymorphism calling without or with a reference genome.

PO0075: Methods: High-throughput Methods

Tilling By Target Capture Sequencing (TbyTCS) to Improve Soybean Seed Composition Traits

Zhou Zhou¹, Naoufal Lakhssassi², Mallory A. Cullen², Abdelhalim El Baz², Shiming Liu², Leandro Neves³ and Khalid Meksem¹, (1)Southern Illinois University at Carbondale, Carbondale, IL, (2)Department of Plant Soil and Agricultural Systems, SIUC, Carbondale, IL, (3)RAPiD Genomics, Gainesville, FL

Chemical mutagenesis emerges as a Genetically Modified (GM)-free strategy to produce large-scale soybean mutants for economically important traits. Reverse genetics approaches have been widely applied in crop species to detect induced mutations in target genes, however, most of them present low efficiency to screen large soybean mutant populations. Here we develop a high-throughput TILLING by Target Capture Sequencing (TbyTCS) technology coupled with universal bioinformatic tools to identify population-wide mutations in soybeans. Because of the robustness of single nucleotide polymorphisms (SNPs) calling, this novel technology ensures high-quality yield of true mutations while removing the majority of false positives. Four Ethyl methanesulfonate (EMS) mutagenized populations (over 4000 mutant families) have been screened for the presence of induced mutations on targeted genes. The mutation types and effects have been characterized for a total of 138 soybean genes in soybean seed composition, disease resistance, and other quality traits. By employing TbyTCS, we discovered novel sources of soybean oil traits as well as protein and carbohydrate traits. EMS-induced mutations at all 19 genes within fatty acid, protein, and carbohydrate biosynthetic pathway including the *GmKASIIA/B*, *GmSACPD-C/D*, *GmFAD2-1A/1B*, *GmFAD3A/B/C*, *GmSus*, *GmGy*, and *GmCG* have been identified. The TbyTCS technology provides an unprecedented platform for highly effective screening polyploidy mutant populations and gene functional analysis. The obtained soybean mutants in this study can be used in subsequent soybean breeding for improved seed composition traits.

PE0076: Methods: High-throughput Methods

Resolving the Phylogeny of Limnephilidae (Insecta: Trichoptera) Using High-Throughput Targeted Enrichment Data and Bioinformatics to Determine the Origins of Unique Life History Traits

Kyle Rawlinson, Brigham Young University, American Fork, UT

The caddisfly family, Limnephilidae (Insecta: Trichoptera), is among the most diverse families within Trichoptera, both in terms of number of species and variety of life history traits. Using high through-put targeted enrichment data to construct the first comprehensive phylogenetic tree for Limnephilidae based on molecular data. The phylogeny will test the following hypotheses: (1) Lineages within Limnephilidae have transitioned from fast to slow moving water and that this movement results in physiological and behavioral changes regarding case making and feeding strategies, and (2) that drought tolerance, including terrestrial and semi-terrestrial larval stages, has arisen within Limnephilidae multiple times, independently.

PO0077: Methods: High-throughput Methods

Majority of Biomass Variation Explained by Drone Images One Day Prior to Harvesting

Zhou Tang, Dept. of Crop and Soil Science, Washington State University, Pullman, WA

Alfalfa is the most widely cultivated forage legume, with approximately 30 million hectares planted worldwide. Genetic improvements in alfalfa have been highly successful in developing cultivars specialized in traits related to winter hardiness and disease resistance. However, genetic improvements have been limited for other economically important traits such as biomass. One of the major bottlenecks for artificial selection is the phenotyping burden for biomass. In this study, we employed two fields to pave a path to overcome the phenotyping burden by using drone images. The first field was used to develop prediction model and the second field to validate the prediction. The first field had 269 alfalfa lines with three replicates. Each replicate was accompanied by 59 plots as checks using a common variety. Three cuttings were harvested and measured for biomass in May, July and September in 2019. The second field had 342 lines with three replicates and each replicate had 75 checks. One cutting was harvested and measured for biomass in September in 2019. The fields were imaged one day prior to harvesting with DJI Phantom 4 pro drone carrying an additional Sentra multispectral camera. Alfalfa plots images were extracted by GRID (<http://zzlab.net/GRID>) software to quantify the vegetative area. The prediction model developed from the first field explained 50~70% (R square) of biomass variation in the second field by four features from drone images, mainly including area. This suggests that high throughput phenotyping with drones could be used for alfalfa biomass selection.

PE0078: Methods: High-throughput Methods

GRID: A Python Package for Aerial High-Throughput Phenotyping

Chunpeng James Chen and Zhiwu Zhang, Dept. of Crop and Soil Science, Washington State University, Pullman, WA

Aerial imagery has the potential to advance high-throughput phenotyping for agricultural field experiments. This potential is currently limited by the difficulties of identifying pixels of interest (POI) and performing their segmentation due to the required intensive manual operations. We developed a Python package, GRID (GReenfield Image Decoder), to overcome this limitation. With pixelwise K-means cluster analysis, GRID users can specify the number of clusters and choose the clusters representing POI. Image grid patterns are automatically recognized by the POI distribution. The local optima of POI are initialized as the plot centers, which can also be manually modified for deletion, addition, and relocation. The segmentation of POI around the plot centers is initialized by automated, intelligent agents to define plot boundaries. A plot intelligent agent negotiates with neighboring agents based on plot size and POI distributions. The negotiation can be refined by weighting more on either plot size or POI density. All adjustments are operated in a graphical user interface with real-time previews of outcomes so that users can incorporate their knowledge of the field site. The final results are saved in text and image files. The text files include plot rows and columns, plot size, and total plot POI. The image files include displays of clusters, POI, and segments. With GRID, users are completely liberated from the labor-intensive task of manually drawing plot lines or polygons. The supervised automation with GRID is expected to enhance the efficiency of agricultural experiments.

Availability: The GRID executable file, user manual, tutorials, and example datasets are freely available at <http://zzlab.net/GRID>.

Contact: chun-peng.chen@wsu.edu

Keywords: Field phenotyping, High-throughput, Aerial image

PO0079: Methods: High-throughput Methods

Ultra High Throughput High Quality Genotypes Using the Applied Biosystems Eureka Genotyping Platform

Victor V. Missirian, John D. Curry, Mohini A. Patil and Ali Pirani, Thermo Fisher Scientific, Santa Clara, CA

The Applied Biosystems™ Eureka™ Genotyping Solution is a low cost, ultra-high throughput (20k samples/week) targeted genotyping by sequencing platform that supports the detection of tens to thousands of genetic markers (SNPs and insertions/deletions). It has been successfully used for a variety of applications (parentage, sex validation, genomic and trait evaluations) both in crops and animals.

The Eureka™ assay is highly resilient to variability in sample composition, including the presence of protein or other cellular debris. The robust nature of the assay enables consistent delivery of high quality genotypes across a wide range of sample types.

We demonstrate that the Eureka™ Genotyping platform consistently delivers high performance (high sample pass rate, call rate, and concordance to known genotypes) across multiple sample types including extracted DNA, dried blood spots, and whole cell lysates from animals. Eureka Genotyping Solution also enables genotypes in 48 hours and high throughput processing to meet quick turnaround demands of the breeding industry.

For Research Use Only. Not for use in diagnostic procedures.

PE0080: Methods: High-throughput Methods

Phenospex Imaging

Grégoire Hummel, Phenospex, Heerlen, Netherlands, Stefan Schwartz, Phenospex, Netherlands and Claude Juriens, ecoRobotix SA, Switzerland

There are more and more autonomous robots available on the market, for different application in agriculture. Those robots seem to be ideal tools to carry sensors to phenotype plots and plants in the field. The robots are light, flexible, cost efficient and fully autonomous, hence an ideal tool to carry sensors for the purpose of plant phenotyping. In this talk we present insights and data of an autonomous robot, which was equipped with a multispectral 3D laser scanner to automate routine application of plant breeders in the field.

PO0081: Methods: Markers

GenoSim: A User-Friendly Simulation Tool for Sequence Reads and SNP Array Genotyping Data in Polyploid Species

Yanlin Liao, Wageningen University and Research, Wageningen, Netherlands

Most recent linkage studies in polyploid crops are based on SNP array data. Genotype by sequencing (GBS) methods are gaining in popularity since they are more flexible and potentially cheaper. Genotype calling in polyploids is challenging due to multiple heterozygous classes that need to be discriminated, requiring high sequencing depth. Here we present GenoSim, a user-friendly software package for the simulation of data from sequencing or SNP arrays. The genotypes can be simulated to come from populations of arbitrary pedigree, size,

ploidy level and mode of inheritance. Based on these simulated genotypes, GBS read counts or array intensities can be simulated with multiple sources of variation, to closely mimic real-life results for various crop species and genotyping technologies. The simulation of genotyping data can be useful to i) study the complexity of genotyping data by modelling the main sources of variation; ii) develop and test genotype calling software or any other software that uses SNP array intensities or read counts as input; iii) study the effect of disturbances in the genotyping data on downstream applications including genetic linkage mapping, QTL mapping and GWAS analyses for polyploids. We suggest parameters to use for approximating the data obtained from SNP array and bait capture sequence reads of potato, chrysanthemum and alstroemeria.

PE0082: Methods: Markers

Mapping Gene Presence-Absence in a Recent Complex Polyploid Crop Genome

Julian Gabur¹, Harmeet Singh Chawla¹, Andreas von Tiedemann², Rod Snowdon¹ and Christian Obermeier¹, (1)Department of Plant Breeding, Justus Liebig University, Giessen, Germany, (2)Department of Crop Sciences, Plant Pathology and Crop Protection Division, Georg August University, Goettingen, Germany

Gene presence absence variation (gene PAV) is known to occur in many gene families of important crop species. The extent and influence of gene PAV on quantitatively inherited agronomic traits is largely unknown. *Brassica napus* is a recent allotetraploid crop genome that has undergone many events of gene loss and speciation and has a strongly rearranged genome. We investigated the association of gene PAV with resistance of oilseed rape (canola) to a fungal pathogen *Verticillium longisporum*, as an example for a complex, quantitatively inherited resistance. Genomic variation was assessed in QTL intervals based on the Brassica Illumina 60k SNP array of bi-parental and multi-parental mapping populations and compared with genome-wide resequencing data of parental lines. In addition, an exon capture array was specially designed to investigate the extent of gene PAV for genes within confidence intervals of quantitative trait loci (QTL). Our results provide insights into the prominent role of gene PAV in quantitatively inherited disease resistance. These findings will improve future breeding efforts on *Brassica napus* and other closely related species.

PO0083: Methods: Markers

Joint Haplotype and Single SNP Analyses Improved Accuracy of Genomic Prediction

Cheng Tan¹, **Zuoxiang Liang²**, Dzianis Prakapenka² and Yang Da², (1)National Engineering Research Center for Breeding Swine Industry, College of Animal Science, South China Agricultural University, Guangzhou, China, (2)Department of Animal Science, University of Minnesota, Saint Paul, MN

We evaluated the accuracy of prediction accuracy using haplotypes and SNPs in the mixed model relative to the prediction accuracy of SNPs. The study population consisted of 7623 individuals with 488,146 SNPs covering all human autosomes from the 2015 version of the Framingham Heart Study data. After quality control, 328,513 SNPs were phased and imputed using BEAGLE (version 5.1). Imputed haplotypes were then divided into blocks with 2-30 SNPs per block. Each block was treated as a locus and each haplotype was treated as an allele. A multi-allelic haplotype model was used to include haplotypes in the prediction model. A 10-fold validation study was conducted to evaluate the prediction accuracy of the prediction model with haplotypes and SNPs relative to the prediction accuracy with SNPs only. The cholesterol phenotypes were analyzed, high density cholesterol, low density cholesterol, and total cholesterol. The results showed that the joint haplotype-SNP model had higher prediction accuracy than the SNP-only model for all three phenotypes. Low density cholesterol had the most improvement in prediction accuracy, 12.46% more accurate than the SNP-only model, followed by high density cholesterol (5.16%) and total cholesterol (3.64%). In domestic animals, genomic selection typically use 50,000-80,000 SNPs. To evaluate feasibility of haplotype analysis using SNPs available in domestic animals, we analyzed the same prediction models using 328,513 SNPs and 82,128 SNPs. The SNP density of 82,128 SNPs was similar to the current SNP density of 80,000 in U.S. dairy genomic selection, and were divided into 0.5 Mb blocks. The results showed that the 82,128 SNPs had similar prediction accuracies as 328,513 SNPs. These results are encouraging for improving prediction accuracy in domestic animals using SNPs currently available.

PE0084: Methods: Markers

Holistic Genotyping of Amplicon Panels with SPAdes and Blastn

Charles F. Crane, USDA-ARS and Department of Botany and Plant Pathology, Purdue University, West Lafayette, IN and Yan M. Crane, USDA-ARS and Department of Entomology, Purdue University, West Lafayette, IN

Although current-generation sequencing has allowed the recognition of myriad SNPs and small indels for use in genetic mapping and association analysis, few studies have investigated multi-SNP markers. Yet such markers might be more accurate and less prone to missing data than single SNPs, depending on the frequency of base-substitution, base-skipping, base-insertion, and run-miscount errors in reads. Failure to detect a particular amplicon allele would affect single-SNP and multiSNP markers equally. MultiSNP markers can exist as many more alleles than single SNPs, allowing all copies to be detected as distinct alleles in polyploids and population samples and greatly simplifying the estimation of allelic dosage and phasing. We propose and test a simple protocol to identify and detect multiSNP genotypes in amplicon panels, where the primer sequence is known but the intervening sequence is not known a priori. We stringently assemble the sequenced, quality-filtered amplicons with SPAdes to generate contigs, which represent all the alleles plus many combinations of sequencing errors. We then align each read to the contigs with blastn and assign the read to the highest-bitscoring contig. We count the number of times each contig is hit and call a genotype from the most frequently hit contigs on the basis of the ploidy and the ratios of counts to one another. We will present simulation results from a diploid, a triploid, and a 14-ploid that represents the alleles in a mapping population derived from heterozygous hexaploid x heterozygous octoploid parents. The simulations will show the response of the called genotypes to sequencing errors and depth of read coverage. We will finish with a real example involving 52 amplified loci in 384 individuals in eight strains of Hessian fly (*Mayetiola destructor*).

PO0085: Methods: Markers

Modeling of QTL Genomic Similarity and Prioritization of Informative SNP Using Fst Scores

Ashley S. Ling¹, Samuel E. Aggrey² and Romdhane Rekaya¹, (1)University of Georgia, Department of Animal and Dairy Science, Athens, GA, (2)University of Georgia, Athens, GA

Use of genomic information has improved selection response in plants and animals. However, little gain in accuracy has been observed beyond low- (wheat) or medium-density (cattle) SNP panels. This has largely been attributed to adequate coverage of the genome by lower density SNP panels, depending on the linkage disequilibrium structure within the population. However, with high-density or sequence data, it is expected that many variants included in the association model or in the calculation of the genomic relationship matrix are not in linkage with QTL. F_{ST} is one population statistic that has been proposed for detection of SNPs segregating with QTL. A simulation was carried out to evaluate the impact of SNPs not in linkage with segregating QTL in prediction accuracy and its ability to distinguish between linked and unlinked SNPs. A 30-chromosome genome with 777k SNPs was simulated. A trait of moderate heritability (0.4) was generated and was assumed to be controlled by 200 QTL located only on 2 chromosomes. Accuracy using all SNPs, linked SNPs (G2), unlinked SNPs, and pedigree was 0.68, 0.92, 0.18, and 0.37, respectively. When SNPs were prioritized using F_{ST} , the top 1k, 10k, 20k, 30k, 40k, and 50k markers yielded accuracies of 0.82, 0.93, 0.92, 0.90, 0.88, and 0.87 and overlapped with the G2 subset at the rate of 100.00, 99.65, 90.32, 74.00, 61.46, and 52.53%, respectively. These results indicate that unlinked SNPs have limited predictive ability to model Mendelian sampling of QTL genomic similarity and that F_{ST} is a useful marker prioritization tool.

PE0086: Methods: Markers

Finding SSRs from WGS Data

Daniel Hansen Lewis, Brigham Young University, Provo, UT

SNPs (Single Nucleotide Polymorphisms) and SSRs (Simple Sequence Repeats) are common polymorphic markers in populations. SNPs have risen in popularity and many tools exist to identify them; however, finding SSRs throughout a population requires running DNA on a gel which is expensive and time consuming. We have developed a Python program, SSR Genotyper, that identifies SSR alleles in a population from WGS data. SSR Genotyper requires a reference FASTA file composed of sequences with an SSR and some upstream and

downstream flanking regions to assist with mapping. It also uses SAM files from accessions that have been mapped this reference file. It outputs a table showing the SSR alleles for each accessions and marker location.

PO0087: Methods: Markers

Automated and Scalable Machine Learning for Whole-Genome Prediction in Plant Breeding

Yalda Zare¹, David Larson² and Paul Skroch², (1)Benson Hill, St Louis, MO, (2)Benson Hill, St. Louis, MO

Automated machine learning (AutoML) is the process of automating the end-to-end process of applying machine learning to real-world problems. Among tasks that are included in this process are cleaning and processing phenotypic and genomic data, selection of genomic features (SNPs), searching for appropriate models and optimization of model hyperparameters. Many of these steps are often beyond the abilities of non-experts. Our proprietary platform CropOS[®] automates these processes to facilitate genomic evaluation and selection of elite crop lines and accelerate the breeding cycle.

We tested the H2O library, an AutoML tool implemented in Python, which trains a large selection of candidate models based on selected maximum runtime or maximum number of models. The candidate algorithms include: Default Random Forest (DRF), Extremely Randomized Forest (XRT), Generalized Linear Models (GLM), XGBoost, Gradient Boosting Machines (GBM), Deep Neural Nets (DeepLearning) and two Stacked Ensemble models. “Stacked Ensemble” models involve training a second-level machine learning algorithm called a “metalearner” to find the optimal combination of the base learners.

Three economically important traits (yield, protein and oil) in a public soybean dataset (13,000 germplasms and 2,000 selected SNPs) were tested. In less than 1 hour, six models were trained for each trait. For yield and oil, Stacked Ensemble was the best model and resulted in higher correlation between predicted and observed phenotypes in the test set compared to each constituent model (0.77 and 0.87, respectively). For protein, XGBoost performed slightly better than Stacked Ensemble (correlation of 0.688 vs. 0.683).

We were able to efficiently run multiple models using the H2O package. H2O automates the manual and tedious step of hyperparameter tuning. Stacked Ensemble models resulted in 1-4% increase in predictive ability compared to Ridge Regression as a baseline model.

PE0088: Methods: Other Genome Methodology

Development of a CRISPR/Cas9 Large DNA Fragment Targeting Technique for Plant Genomes

Carine Satgé¹, Caroline Callot¹, **Stephane Cautet**², William Marande¹, Pauline Duriez³, Stéphane Muñoz⁴, Sonia Vautrin¹ and Margaux-Alison Fustier¹, (1)French Plant Genomic Center CNRGV - INRA, Castanet-Tolosan, France, (2)French Plant Genomic Center CNRGV - INRAE, Castanet-Tolosan, France, (3)Laboratoire des Interactions Plantes Micro-organismes (LIPM) - INRA/CNRS, Castanet-Tolosan, France, (4)Laboratoire des Interactions Plantes Micro-organismes (LIPM) - INRAE/CNRS/Université de Toulouse, Castanet-Tolosan, France

To accelerate plant breeding, high quality assemblies are essential. They help to understand genome structure and to identify genes involved in agronomic traits including resistance to various environmental stresses, resistance to pathogens and high yield. However, the exploration of plant genomes remains challenging due to the complexity of plant genomes in terms of size, repetitive elements content and various levels of ploidy. Moreover, because of a high intra-species variability, a quality reference sequence is not enough to obtain a precise and reliable information of a genomic region linked to a trait of interest in a specific genotype.

New strategies for efficiently targeting large regions of interest in complex genomes are really needed to be able to link a phenotype to a genotype.

Here, we investigate the potential of the CRISPR/Cas9 system to target a 120 kbp genomic region of interest from a complex genome, the sunflower *Helianthus annuus*. We improved and adapted the first steps of the CATCH method (Cas9-Assisted Targeting of CHromosomal segments as described by Jiang et al., 2015). Then, we sequenced the targeted region with long reads sequencing approach coupled to the PacBio low input protocol. This strategy

allowed the enrichment of the genomic region with high quality assembly. Thus, we propose a CRISPR/Cas9 based method amplification-free, with a simplified bioinformatical pipeline and a potential for multiplexing to sequence large genomic region of interest from plant genomes.

PO0089: Methods: Other Genome Methodology

DeepTE: A Computational Method for De Novo Classification of Transposons with Convolutional Neural Network

Haidong Yan, School of Plant and Environmental Sciences, Virginia Tech, USA, Blacksburg, VA, Song Li, Virginia Tech, Blacksburg, VA and Aureliano Bombarely, Virginia Polytechnic Institute and State University, Department of Horticulture, Blacksburg, VA; Department of Bioscience University of Milan, Milan, VA, Italy

Transposable elements (TEs) classification is an essential step to decode their roles in genome evolution. With a large number of genomes from non-model species becoming available, accurate and efficient TE classification has emerged as a new challenge in genomic sequence analysis.

We developed a novel tool, DeepTE, which classifies unknown TEs using convolutional neural network. DeepTE transferred sequences into input vectors based on of k-mer counts. A tree structured classification process was used where eight models were trained to classify TEs into super families and individual TE order. DeepTE also detected domains inside TEs to correct false classification. An additional model was trained to distinguish between non-TEs and TEs in plants. Given exclusive TEs of different species types, DeepTE classified seven orders, and 15, 24, and 16 super families in plants, metazoans, and fungi, respectively.

DeepTE outperformed other existing tools for TE classification in our benchmarking experiments. This tool successfully leverages convolutional neural network for TE classification, assisting to precisely identify and annotate TEs in newly sequenced eukaryotic genomes.

PE0090: Methods: Other Genome Methodology

Bionano Hybrid Scaffold Assemblies Provide High Contiguity and Accuracy

Andy Wing Chun Pang, Joyce Lee, Jian Wang, Karl Hong, Ernest Lam, Thomas Anantharaman, Mark Oldakowski and Alex R. Hastie, Bionano Genomics, San Diego, CA

With the advancement of new technologies, high quality assemblies of novel genomes have gained momentum in recent years. To disambiguate homologous regions in these novel genomes, long reads and linked reads are used to assemble contigs. Scaffolding these sequences into chromosomal-arm or full chromosome length can only be accomplished using Bionano Genomics optical mapping or one of the Hi-C based methods. In comparing the assemblies of these scaffolding technologies, we demonstrate that Bionano can correct sequence and orientation errors generated by other technologies, while providing superior contiguity.

Chromosome assemblies can often be achieved by using Bionano's new DLS chemistry. Bionano Genomics, with its physically intact molecules that have N50 length on average >250kbp, is unique and can generate megabase-long contiguous assemblies with the well understood overlap-layout-consensus algorithms. These assemblies are then used to scaffold sequences into chromosome or chromosome arm length assemblies by the Bionano Hybrid Scaffold pipeline. Alternatively, Hi-C based methods leverage crosslinking of DNA that is in close-proximity *in vivo* through chromatin folding, which is then sequenced using short read sequencing. Since the long-range interaction in Hi-C is based on cells at different stages of dynamic biological connections and is encoded by short reads, significant inference is required to reconstruct the interaction information.

We present here a few case studies of scaffolding plant and animal genomes using Hi-C and Bionano maps. We examine some of the challenges and suggest an effective workflow in generating high-quality reference-graded assemblies.

PO0091: Methods: Other Genome Methodology

Corrected Pacbio Iso-Seq CCS to Improve Genome Gene Annotation

Shengqiang Shu¹, Mike Frizzell², Jane Grimwood², David M. Goodstein¹, Jeremy Schmutz² and Daniel Rokhsar¹, (1)DOE Joint Genome Institute, Walnut Creek, CA, (2)HudsonAlpha Institute for Biotechnology, Huntsville, AL

PacBio Iso-Seq circular consensus sequences (CCS) from either Sequel system or Sequel II system are short-read corrected and then clustered after alignment against their respective reference genome. These clustered CCSs can often be treated in gene-calling pipelines as equivalent to full length cDNAs, when supported by sequencing metadata (e.g., presence of 5', 3' and poly(A) tail) and alignment quality, and can significantly improve protein coding gene structure quality. Using putative full length transcript status of clustered CCSs and cluster size, false positive alternative transcripts from short read assemblies can be filtered out and gene locus boundaries can be defined with higher confidence. We have used corrected and clustered CCSs in our both versions of our plant genome gene annotation pipelines (GMI - gene model improvement and IGC - integrated gene call), and find that the resulting gene sets have higher BUSCO score (i.e., fewer core plant genes are missed) and more likely correct UTRs.

PE0092: Methods: Other Genome Methodology

False Gene Losses Corrected By VGP Assembly

Juwan Kim¹, Chul Lee¹, Byung June Ko², Arang Rhie³, Heebal Kim^{1,2}, Erich D. Jarvis⁴ and VGP assembly group, (1)Interdisciplinary Program in Bioinformatics, Seoul National University, Seoul, South Korea, (2)Department of Agricultural Biotechnology and Research Institute of Agriculture and Life Sciences, Seoul National University, Seoul, South Korea, (3)NHGRI, NIH, Bethesda, MD, (4)The Rockefeller University, New York, NY

A reference assembly of a species should fully reflect the structure of genes within its genome. But previous assemblies may lead to incorrect disruption of the reading frame of protein coding genes due to assembly errors. Here, we compared high-quality chromosome-level reference assemblies generated by Vertebrate Genomes Project (VGP) with previous ones to identify genes with incorrectly lost reading frames on previous assemblies. Based on the projection of annotations by Comparative Annotation Toolkit (CAT), we detected numerous false gene losses on four species: zebra finch, platypus, tiger fugu and climbing perch which have both VGP and previous assemblies. We classified the false gene losses in previous assemblies into one of the eight types: totally missing gene, gene mapped below 50% of coding sequences, fragmented gene, intra-scaffold split gene, premature stop codon, frameshift, gap (N) in coding sequences, and intron exon junction disruption. Comparing the amount of false gene losses in sanger and pacbio-based assembly, we confirmed that there were greater number of false gene losses in sanger-based assemblies which implies the relationship between sequencing platform and the amount of false gene losses. We focused on the types of genes that reported as premature stop codons, frameshifts, or intron exon junction disruption, which can be considered as major artifacts of sequencing errors. In particular, the genes with frameshifts, the most frequently reported, were compared with orthologues of other species to compare the accuracy of annotations of VGP with the previous assemblies. Our study shows the importance of high quality reference assembly for accurate annotation.

PO0093: Methods: Other Genome Methodology

False Duplications and Gains of Genes in Previous Genome Assembly: Revealed by High-Quality VGP Assembly

Byung June Ko¹, Chul Lee², Juwan Kim², Erich D. Jarvis³, Heebal Kim^{1,4} and VGP assembly group, (1)Department of Agricultural Biotechnology and Research Institute of Agriculture and Life Sciences, Seoul National University, Seoul, South Korea, (2)Interdisciplinary Program in Bioinformatics, Seoul National University, Seoul, South Korea, (3)The Rockefeller University, New York, NY, (4)C&K Genomics, Seoul, South Korea

To identify and to minimize assembly errors is the one of the key processes to construct reliable reference genome. Recently, high-quality reference genomes have been produced by Vertebrate Genome Project (VGP) with long reads, and make it possible to observe assembly errors in the previous reference genome. In this research, we investigated some duplicated regions in previous reference genomes assumed as phasing errors and they may cause some false gains of genes and exons. Here, we analyzed genomes with both of VGP and previous assemblies and annotations by using CACTUS genome-wide alignments: zebra finch (*Taeniopygia guttata*), platypus (*Ornithorhynchus anatinus*) and torafugu (*Takifugu rubripes*). We identified duplicated sequences in homologous regions among VGP and previous assemblies. In the case of zebra finch, phased sequences of allele-relation were aligned together for identifying false duplicated loci. We found 48,004 homologous loci containing false duplication in the previous Sanger assembly of zebra finch. Out of these loci, 280 loci included the 100 of false gene gain cases on previous assembly which were not found on VGP assembly. On the other hands, we identified 218 genes with exon duplications only in the previous genome. We believe this tendency of false duplications genomes give opportunities to update assembly processes and to understand genomic evolution including whole genome duplications such as, polyploidy.

PE0094: Methods: Other Genome Methodology

How Genomic Relationships Impact the Prediction Accuracy of Breeding Values

Beatriz Cuyabano and Cedric Gondro, Michigan State University, East Lansing, MI

Genomic models incorporating dense marker information are widely used to predict breeding values in order to select animals and plants in breeding programs. Among other factors, the accuracy of genomic prediction depends on the heritability of the trait and the structure of the genomic relationships between genotyped individuals in the reference and test populations. To understand these relationships, we used genotypic data from populations with different genetic structures (human, sheep, cattle and pig) combined with real and simulated phenotypes. A resampling strategy was used to split the data into reference and test subsets which were then used to estimate heritabilities and evaluate prediction accuracies. We observed strong negative correlations between heritability estimates in the reference populations and prediction accuracies in the test populations, which was consistent across different populations and sample sizes. This is an interesting resampling property of the relationship between heritability and prediction accuracy which arises in finite populations since the total amount of variance is fixed. We then derived a new equation to predict the accuracy of genomic prediction conditioned on the genomic relationships between reference and test populations that is more accurate than methods based on linkage disequilibrium. Results from this work can be used to select specific individuals for the reference population that improves the prediction accuracy of a test population.

PO0095: Methods: Other Genome Methodology

Combining Various Genomic Strategies With New Technologies to Decipher the Complex Structure of Plant Genomes

Stephane Cauet¹, Nathalie Rodde², Carine Satgé², Margaux-Alison Fustier², Caroline Callot², William Marande², Sandrine Arribat², Arnaud Bellec¹ and Sonia Vautrin², (1)French Plant Genomic Center CNRGV - INRAE, Castanet-Tolosan, France, (2)French Plant Genomic Center CNRGV - INRA, Castanet-Tolosan, France

Among living organisms, plants display a high level of genomes complexity due to their large size, variations in ploidy levels and high percentage and variability of repetitive elements. Despite the Next Generation Sequencing revolution including the recent long read technologies, it remains challenging to obtain high quality assemblies at the genome scale. In order to be efficient, when addressing a scientific question, it is important to choose the relevant strategy according to with the raised topic: exhaustive information on whole genome is not always required while reliable and quality information of the region of interest is crucial and necessary. A reliable sequence information linked to a trait of interest in specific genotypes is essential to understand the role of a genomic region in a phenotype.

The French Plant Genomic Resources Center (CNRGV) provides various innovative and efficient genomic tools to better characterize plant biodiversity and understand how plants adapt to their environment through the analysis of

their genomes and the intra/inter-species variability. We develop several strategies combining large fragment genomic DNA libraries, CRISPR-CATCH targeting strategy and optical mapping technology combined with long reads sequencing technologies to obtain very high quality sequence. The complementarity of these strategies allows the production of reliable sequence information, which is essential to link a genotype to a phenotype.

PE0096: Methods: Other Genome Methodology

Identification of *C. clementine* Genome Assembly Anomalies Through WGA of Single Pollen Grains and High-Density SNP Array

Sergio Pietro Ferrante, Yoko Eck, Zhenyu Jia and Mikeal Roose, University of California, Riverside, CA

To reveal the haplotypes of the donor plant, DNA from single pollen grains was whole genome amplified (WGA) and processed on the Axiom™ Citrus56 Array (Affymetrix, Inc.), which includes 57,933 autosomal and 500 chloroplast SNPs. 221 WGA of single pollen grains from 39 diverse diploid accessions were successfully genotyped for at least 97% of SNPs (success rate ~71% of samples). For each genotype 5 genotyped pollen grains were used to reconstruct the donor's haplotypes. Anomalies in the haplotype reconstruction were identified in the same regions of several chromosomes (CHRs) among many of the 39 genotypes. Along the CHRs of each genotype, anomalies were revealed as an abnormal number of crossovers, at the same genome position, among the 5 pollen grains. If marker density is high relative to the frequency of crossovers, the probability that two gametes have break points between the same pair of adjacent markers is expected to be close to zero. The length of the anomalies (Mb) along each CHR, varied according to the number of heterozygous SNPs present in that region in each genotype. The detected anomalies were compared to the deviant genome sequence positions identified by linkage mapping (JoinMap 5.0) in several segregating populations of citrus genotyped with the Axiom™ Citrus56 Array. The deviant genome sequence positions identified by linkage mapping, overlapped with anomalous chromosomal regions detected through WGA of single pollen grains, highlighting the presence of possible genome assembly errors in the Clementine reference genome, or that several rearrangements are common in other citrus varieties.

PO0097: Methods: Other Genome Methodology

Tool Development Via Modelling for Ambiguous SNP Calls from Amplicon Derived Data in Allotetraploid Peanut

Roshan Kulkarni¹, Karin Dorman¹, Hongxu Wang¹, Andrew Wilkey², Nathan T. Weeks², Mark D. Burow³ and Steven Cannon², (1)Iowa State University, Ames, IA, (2)USDA-ARS-CICGRU, Ames, IA, (3)Texas A&M AgriLife Research, Lubbock, and Department of Plant and Soil Science, Texas Tech University, Lubbock, TX

Cultivated peanut is an allotetraploid crop with highly similar A and B sub-genomes coupled with large genome size of around 2.7 Gbps. Accurate genotyping of allotetraploid peanut is challenging due to alignment ambiguities caused by homoeology leading to an excess of heterozygous calls. In this study we propose an allotetraploid specific method that carefully assesses the strength of the A and B alignments to estimate the genotype of a sequenced individual at a single locus in a homoeologous region. The paired end reads derived from targeted resequencing were merged using a custom read merger and then aligned to the A and B genome targets separately using the Burrows Wheeler Aligner. The log likelihoods of each alignment were computed using quality scores and assuming independence of nucleotides and equally likely substitutions. The alignments and their posterior probabilities were used to estimate a flexible sequencing error model. The recalibrated error model is then used to estimate, via Monte Carlo sampling, the likelihood of all n reads aligned to the SNP region given the genotype. Then, the genotype that maximizes the likelihood is reported as the estimated genotype. In providing this tool, we hope to benefit plant breeding programs by genotyping allotetraploids with greater accuracy and thereby better revealing the true variations among genotypes.

PE0098: Methods: Other Genome Methodology

Y Chromosome Expansion Slowdown in *Silene latifolia*

Jan Safar¹, Wojciech Jesionek², Zdenek Kubat³, Janka Puterová⁴ and Roman Hobza³, (1)Institute of Experimental Botany, Olomouc, Czech Republic, (2)Institute of Biophysics ASCR, Brno, Czech Republic, (3)Institute of Biophysics ASCR vvi, Brno, Czech Republic, (4)Institute of Biophysics, Brno, Czech Republic

Sex chromosomes represent a specific genomic region(s) with locally suppressed recombination. As a consequence, repetitive DNA accumulates on the Y (W) chromosomes. To assess the size dynamics of the Y chromosome, we studied intraspecific genome size variation and genome composition of male and female individuals in a dioecious plant *Silene latifolia*, a well-established model for sex-chromosomes evolution. The rise and fall of the Y chromosome was demonstrated in animals but plants often possess the large young Y chromosome that is thought to have expanded recently. We evaluated the impact of individual repetitive DNA fractions on genome size and the Y chromosome expansion in selected *S. latifolia* populations. We tested whether the most active repeats (those biased in populations with large genomes) have also been affecting recent Y chromosome structural changes. We show that epigenetic mechanisms play a major role in regulation of Y chromosome dynamics.

PO0099: Methods: Other Genome Methodology

A Method to Map Structural Variation

Luca Comai¹, Kirk Amundson², Benny Ordonez², Xin Zhao³, Guilherme Braz⁴, Jiming Jiang⁴ and Isabelle M. Henry², (1)Plant Biology and Genome Center, UC Davis, Davis, CA, (2)UC Davis, Davis, CA, (3)University of California Davis, Davis, CA, (4)Michigan State University, East Lansing, MI

Genomic Structural variation (SV) is common and has a profound effect on the phenotype of individuals. Chromosomal translocations are an important type of SV, but also a common spurious effect of defective genome assemblies. In theory, they can be identified by detecting unexpected linkage, by sequencing unexpected junctions, and by fluorescent in situ hybridization (FISH). In practical terms, however, identifying and characterizing a translocation is not simple, particularly in the absence of prior evidence pointing to its location. We study structural genomic variation and developed a method to identify translocations that is based on low-pass Illumina sequencing of related segregating individuals. By comparing sequence coverage in individuals such as siblings, we can cluster genomic dosage states at SV loci. We then detect unexpected linkage through covariance analysis of these copy number variable loci. We demonstrate this method using a population of dihaploid individuals produced by haploid induction crosses in cultivated, autotetraploid potato (*Solanum tuberosum*). In this background, we document frequent nonreciprocal chromosomal translocations as well as possible assembly errors. The deleterious effect of unbalanced translocations in potato is likely buffered by polyploidy. Furthermore, purging of deleterious SV by meiosis is hindered by clonal propagation.

PE0100: Methods: Other Genome Methodology

De novo* Genome Assembly based on the Integration of Illumina and Nabsys Data Unraveled Important Bio-Control Features of the Endophyte *Lysinibacillus fusiformis

Alessandro Passera¹, Paola Casati¹, John Oliver², Emanuela Consentino³, Michael Kaiser⁴, Luca Marcolungo³, Jay M. Sage², Jennifer R. Davis², Barrett Bready², Massimo Delledonne⁵ and **Marzia Rossato³**, (1)University of Milan, Milan, Italy, (2)Nabsys, Providence, RI, (3)University of Verona, VERONA, Italy, (4)Nabsys, na, RI, (5)University of Verona, Verona, Italy

Lysinibacillus genus includes several plant-associated soil bacteria, which can have interesting agricultural and biotechnological applications. However, little is known about this genus especially when considering genomic information.

We have characterized the genome of *Lysinibacillus fusiformis*, strain S4C11, that has some antifungal activities and plant-associated features. A hybrid assembly approach was used to integrate the short-read sequencing data obtained by Illumina technology and the whole genome mapping data generated using the novel approach of High-Definition Electronic genome mapping provided by Nabsys. *De novo* genome assembly based solely on Illumina reads generated a relatively fragmented assembly of 5.07 Mbp in 55 ungapped sequences with N50 length of 584 Kbp. Hybrid assembly, integrating the Illumina-based assembly with the HD electronic maps by Nabsys, allowed to

close/finish the genome assembly in a single circular chromosome of 4.7 Mbp. In addition, Nabsys data discovered the presence of two additional chromosomes/relatively large plasmids (193 kbp and 137 kbp), not recognized by pure-bioinformatic search.

Genome annotation identified 5033 coding sequences and enrichment analysis unraveled pathways of particular interest for endophyte biology. These included genes related to the production and utilization of flagella for movement and chemotaxis, important traits for rhizosphere competence. The production of siderophores and bacteriocins, important for inhibiting the growth of plant pathogens, was also identified, as well as the capability to produce proteins that detoxify ROS, crucial for endophytes to survive inside the plant. Finally, genes of the chitin utilization pathway were also identified, that support the antifungal activity of this strain.

PO0101: Methods: Other Genome Methodology

Comparative Transcriptomics and Metabolite Profiling Unravel the Formation and Biochemical Specialization of Trichomes on Tea Plant for Defense and Defensive Secondary Metabolite Production Integrated By CsMYB184

Jian Zhao, Anhui Agricultural University, Hefei, China

State Key Laboratory of Tea Plant Biology and Utilization, College of Tea and Food Science and Technology, Anhui Agricultural University, Hefei, 230036,

Abstract

Tea plant (*Camellia sinensis*) leaf produces a wide array of secondary metabolites, including catechins, caffeine, and theanine as characteristic bioactive ingredients of teas, which make major contributions to tea flavors and health benefits. Trichomes on tea plant leaves produced many special flavors-contributing products and are regarded as an important quality trait, however, the underlying molecular bases for how tea unicellular trichomes are generated on plant leaf to synthesize specialized metabolites, and what's their physiological function in tea plant are not fully understood. Here, we integrated metabolome and transcriptome analyses on tea trichomes and trichome-removed leaves (TR-leaves) to gain insight into the formation and functions of trichomes as well as their synthesis of specialized metabolites. Metabolite profiling and RNA-Seq data strongly supported the defense functions of trichomes by highly expressing many defensive proteins/enzymes, or synthesizing and accumulating a complex defensive secondary metabolites, some of them like mono- and sesquiterpenes, contributing to the essential role of trichomes in determining tea flavors. Besides producing these characteristic tea secondary metabolites, tea plant trichomes also highly and specifically accumulated more defense metabolites and expressed more defensive genes, such as UV-protective flavonols and peptides, insect-toxic caffeine and proteinase inhibitors, and other volatiles, disease-resistant metabolites and proteins. Tea trichomes also specifically express defense-related germin-like proteins, chitinase, laccase, GST, high-light protective proteins, and LRR protein kinase family genes, which are usually associated with plant defense against herbivore insects and pathogens and abiotic stresses. We functionally characterized CsMYB184, CsGL3, and CsWD40-repeat genes are highly associated with trichome formation and development with molecular and genetic tools. Antisense knockdown of CsMYB184 also reduced expression levels of gene involved in biosynthesis of volatiles, catechin, and caffeine, followed by reduced production of these metabolites in buds and young leaves. Nucleus-localized CsMYB184 activated the promoters of ANR and ANS and a tea plant GL2 homolog CsGL2 genes in a reporter activation assay. Association studies on trichome phenotyping and gene expression in various tea plant germplasm demonstrated a close relation between CsMYB184 expression level and trichome density and lengths. Our findings suggest that tea plants have evolved to use trichomes for defense and adaptation to adversary environments, also by enhanced metabolic capabilities for efficient production of diverse defensive metabolites.

PE0102: Methods: Other Genome Methodology

Population Structure Inference using Phenotype Data

Anthony Greenberg, Bayesic Research, Ithaca, NY

Assigning individuals to populations is important for a variety of applications. Given the abundance of genotype data, multiple widely-used methods are available for population structure inference. Employing these methods, we have learned a great deal about evolutionary history of species. Phenotypic differences among populations are then typically studied given the genotype-based population assignments. However, it would be useful to infer population differentiation using phenotypic data alone. Studying the concordance, or lack thereof, between genotype and phenotype driven assignments can yield insights into local adaptation and identify useful sources of breeding material for crop and livestock improvement. I will describe a Bayesian Gaussian mixture model that uses (possibly replicated) measurements of multiple traits to infer population structure. I will present an R package, MuGaMix, that can perform such inference. Its performance will be assessed on simulated and real data.

PO0103: Methods: Other Genome Methodology

Design of Experiments for Fine-Mapping Quantitative Trait Loci in Livestock Populations

Dörte Wittenburg and Michael Doschoris, Leibniz Institute for Farm Animal Biology, Dummerstorf, Germany

Single nucleotide polymorphisms (SNPs) with significant impact on a trait can be identified with genome-wide association studies. High linkage disequilibrium (LD) among SNPs makes it difficult to identify causative variants correctly. Thus, often target regions instead of single SNPs are reported. Sample size has not only a crucial impact on the precision of parameter estimates, it also ensures that a desired level of statistical power can be reached. We study the design of experiments for fine-mapping quantitative trait loci (QTL) in such a target region.

A multi-locus model allows to identify causative variants simultaneously, to state their positions more precisely and to account for existing dependencies. Based on the commonly applied SNP-BLUP approach, we determine the z-score statistic for locally testing non-zero SNP effects and investigate its distribution under the alternative hypothesis. This quantity employs the theoretical instead of observed dependence between SNPs; it can be set up as a function of paternal and maternal LD for any given population structure.

We simulated multiple paternal half-sib families and considered a target region of 1 Mbp. A bimodal distribution of estimated sample size was observed, particularly if more than two QTL were assumed. The first mode roughly coincided with sample size estimated from single-SNP investigations (e.g. QUANTO). The second mode pointed to inflated sample sizes and could be explained by blocks of varying linkage phases of sires leading to negative correlations between SNPs. In ongoing work, we will verify if the simulated linkage blocks resemble the genome structure in dairy cattle justifying increased sample sizes for fine-mapping experiments.

PE0104: Methods: Other Genome Methodology

BHQ Probe Master Mix, a Versatile Master Mix for qPCR Applications on Agricultural Sample Types, Including AP Testing, CNV Determination and Pathogen Detection

Luke Linz, LGC, Biosearch Technologies, Alexandria, MN

BHQ Probe Master Mix from LGC, Biosearch Technologies, has been established as a robust PCR Master Mix for fluorescent end-point genotyping applications. Here we present data validating its use in qPCR, utilising a range of BHQ-type hydrolysis probes from the Biosearch Technologies portfolio. The data demonstrates high levels of reproducibility and sensitivity in a variety of agrigenomics applications, including adventitious presence (AP), copy number variation (CNV) and pathogen detection in a variety of agricultural sample types. Quantification of genes known to have CNV in common wheat varieties shows high accuracy when compared to endogenous standards when using BHQnova probes, and AP was detected down to 1% in wheat and maize, and low as 0.1% in soybean (compared to Certified Reference Material from LGC Standards) using dual-labelled BHQ probes. In addition, high sensitivity and specificity of BHQ Probe Master Mix with BHQplus probes was demonstrated with the detection of two common plant pathogens from the *Globodera* genus, *G. pallida* (potato cyst nematode) and *G. rostochiensis* (golden nematode). The data presented here clearly shows the use of the BHQ Probe Master Mix in qPCR applications for robust and reproducible detection and quantification of low copy-number and low-frequency targets, from a range of challenging agricultural sample types.

PO0105: Methods: Other Genome Methodology

Runs of Homozygosity and Analysis of Inbreeding Depression

Pattarapol Sumreddee¹, Sajjad Toghiani², Elhamidi Hay³, Samuel E. Aggrey¹ and Romdhane Rekaya⁴,
(1)University of Georgia, Athens, GA, (2)USDA-ARS-PA-Livestock & Range Res. Lab (LARRL), Miles city, MT, (3)USDA-ARS-PA-Livestock & Range Res. Lab (LARRL), Fort Collins, CO, (4)University of Georgia, Department of Animal and Dairy Science, Athens, GA

Assessment of autozygosity across chromosomal segments using runs of homozygosity (**ROH**) has emerged as a valuable tool to estimate inbreeding due to its general flexibility and ability to quantify chromosomal contribution to genome-wide inbreeding. Unfortunately, identifying ROH segments is sensitive to the parameters used during the search process. These parameters are heuristically set, leading to significant variation in the results. The minimum length required to identify an ROH segment has major effect on the estimation of inbreeding. A search algorithm to approximate mutation loads was developed to determine the minimum length of ROH segments. It consists in finding genome segments with significant effect differences in trait means between animals with high and low autozygosity intervals at certain threshold values. The minimum length could be determined as the smallest interval at which a significant signal is detected. The method was tested in a Hereford population genotyped for 30,220 SNPs. Phenotypes recorded for six traits were used for the approximation of mutation load. The estimated minimum length was around 1 Mb for yearling weight (**YW**) and average daily gain (**ADG**), and 4 Mb for birth weight (**BW**) and weaning weight (**WW**). These trait-specific thresholds are the result of trait-dependent effects of homozygosity. The onset of significant inbreeding effects was well aligned with the estimated thresholds, especially for YW and ADG. Our results highlight the importance of accurate estimation of the ROH-based inbreeding and the necessity to consider a trait-specific minimum length threshold for the identification of ROH segments in inbreeding depression analyses.

PE0106: Methods: Other Genome Methodology

Genomic Predictions for Montana Composite Cattle using Metafounders

Sabrina Kluska¹, Yutaka Masuda², Fernando Baldi¹, Joanir P. Eler³, Jose B. S. Ferraz³ and Daniela Lourenço²,
(1)São Paulo State University, Jaboticabal, Brazil, (2)University of Georgia, Athens, GA, (3)University of Sao Paulo/College of Animal Science and Food Engineer/FZEA, Pirassununga, Brazil

Metafounders are pseudo-individual added to the pedigree to consider relationship within and among base populations. The aim of this study was to investigate the impact of using metafounders and random unknown-parent groups (UPG) in genomic evaluations of a composite beef cattle population. Scrotal circumference measurements at 12 months of age (SC12) were available for 55,153 animals. Among those, 1899 were genotyped for 30,000 SNP. Five analyses were performed: traditional pedigree BLUP, pedigree BLUP with UPG (BLUP_UPG), single-step GBLUP (ssGBLUP), ssGBLUP with UPG (ssGBLUP_UPG) and ssGBLUP with metafounders (ssGBLUP_MF). The Montana composite population traces back to 5 different ancestral groups, therefore, 5 UPG or metafounders were used. Validation was done in 759 genotyped animals that had their phenotypes removed from the evaluations. Predictive ability was computed as the correlation between adjusted phenotypes and (G)EBV for validation animals. Inflation was defined as the b1 of the regression of adjusted phenotypes on (G)EBV. The predictive ability was 0.29, 0.29, 0.34, 0.34, and 0.33 for BLUP, BLUP_UPG, ssGBLUP, ssGBLUP_UPG, and ssGBLUP_MF models, respectively. The b1 was 1.04, 1.05, 1.03, 1.03, and 1.07, in the same above mentioned order. The predictive ability increased when genomic information was added to the models (0.05 points). Among all models, ssGBLUP and ssGBLUP with UPG had the least deflated predictions. In this Montana composite beef cattle population, adjusting the genetic base using unknown parent groups or metafounders did not help to increase model performance, possibly because of limited population size.

PO0107: Methods: Other Genome Methodology

Customizable, Fast, and Easy Method for SNP and Target Detection

Emily Jesgarz, MatMaCorp, Lincoln, NE

From identification of causative mutations to the detection of infectious agents, SNP and target detection applications are broadly utilized in numerous industries including healthcare, agriculture, and veterinarian medicine.

A variety of methods, technologies and instrumentation have been incorporated into laboratory, point of care (POC), and point of use (PU) applications. Most of these methods require skilled personnel and common laboratory equipment such as pipettes, centrifuges, and refrigerators. Additionally, rapid development of needed SNP or target based assays are rare and cost prohibitive. We have developed a method called C-SAND (combined sequence amplification and nucleotide detection) which allows rapid custom development of SNP or target detection assays. The SNP or targets are identified using labeled primers on a four-channel fluorescence detection device called Solas 8. Additionally, all the reagents are lyophilized and provided as a two-step method which has eliminated the need for common laboratory equipment. Lab versions of the test that require minimal pipetting are also an option, allowing the user flexibility to quickly execute this kit in the field or lab. For fast evaluation of informative SNPs in the lab and to the rapid deployment of a diagnostic test to detect emerging threats in the field, C-SAND assays offer affordability, flexibility, and ease of use.

PE0108: Methods: Sequencing

A Complete Solution for High-Quality Genome Annotation Using the PacBio Iso-Seq Method

Elizabeth Tseng, Pacific Biosciences, Menlo Park, CA

The PacBio Iso-Seq method produces high-quality, full-length transcripts of up to 10 kb and longer and has been used to annotate many important plant and animal genomes. We describe here the full Iso-Seq ecosystem that enables researchers to achieve high-quality genome annotations. The Iso-Seq Express workflow is a 1-day protocol that requires only 60-300 ng of total RNA and supports multiplexing of different tissues. Sequencing on a single SMRT Cell 8M on the Sequel II System produces up to 4 million full-length reads, sufficient to exhaustively characterize a whole transcriptome on the order of 15,000-17,000 genes with 100,000 or more transcripts. Most importantly, the method is supported by a maturing suite of official and community-developed tools. The SMRT Link Iso-Seq application outputs high-quality (>99% accurate), full-length transcript sequences that can optionally be mapped to a reference genome for a single SMRT Cell worth of data in 6-9 hours. For example, the SQANTI2 tool classifies Iso-Seq transcripts against a reference annotation, filters potential library artifacts, and processes information from both long read-only and short read-based quantification. IsoPhase is a tool for identifying allele-specific isoform expression. Cogent has been used to process Iso-Seq transcripts in a genome-independent manner to assess genome assemblies. Finally, IsoAnnot is an up-and-coming tool for identifying differential isoform expression across different samples. We describe how these tools complement each other and provide guidelines to make the best use out of Iso-Seq data for understanding transcriptomes.

PO0109: Methods: Sequencing

Enzymatic Methyl-Seq: Next Generation Methyloases

Fiona J Stewart¹, V.K. Chaithanya Ponnaluri¹, Brittany S Sexton¹, Louise Williams¹, Bradley W Langhorst¹, Romualdas Vaisvila¹, Yvonne Helbert², Lei Zhang², Timothy Harkins², Kevin McKernan², Eileen T Dimalanta¹ and Theodore B Davis¹, (1)New England Biolabs, Ipswich, MA, (2)Medicinal Genomics, Beverly, MA

DNA methylation is important for gene regulation. The ability to accurately identify 5-methylcytosine (5mC) and 5-hydroxymethylcytosine (5hmC) gives us greater insight into potential gene regulatory mechanisms. Bisulfite sequencing (BS) is traditionally used to detect methylated Cs; however BS does have its drawbacks. DNA is commonly damaged and degraded by the chemical bisulfite reaction, resulting in libraries that demonstrate high GC bias enrichment for methylated regions. To overcome these limitations, we developed an enzymatic approach, NEBNext Enzymatic Methyl-seq (EM-seqTM), for methylation detection that minimizes DNA damage, resulting in longer fragments and minimal GC bias.

Illumina libraries were prepared using bisulfite and EM-seq methods with 50 ng DNA from *Arabidopsis thaliana* and *Cannabis sativa* DNA. Libraries were sequenced using Illumina's NextSeq 500. Reads were aligned using BWAMeth 0.2. and methylation information was extracted using MethyDackel. Total 5mC levels were compared between the sequencing data from EM-seq and WGBS libraries and LCMS (Liquid Chromatography Mass Spectrometry). 5mC levels determined by EM-seq are close to those from LCMS, whereas WGBS results in an overestimation of 5mC. Additionally, EM-seq libraries produce higher quality sequencing metrics such as longer

inserts, lower duplication rates, a higher percentage of mapped reads and less GC bias compared to bisulfite converted libraries. We conclude that EM-seq is superior to WGBS and delivers higher library yields, more accurate methylation information, reduced DNA damage, increased sequencing length, and decreased GC bias.

PE0110: Methods: Sequencing

Beyond Contiguity: Evaluating the Accuracy of *de novo* Genome Assemblies

Sarah B Kingan¹, Zev Kronenberg² and Aaron Wenger², (1)Pacific Biosciences, Menlo Park, CA, (2)PacBio, Menlo Park, CA

HiFi reads (>99% accurate, 15-20 kb) from the PacBio Sequel II System consistently provide complete and contiguous genome assemblies. In addition to completeness and contiguity, accuracy is of critical importance, as assembly errors complicate downstream analysis, particularly by disrupting gene frames. Metrics used to assess assembly accuracy include: 1) in-frame gene count, 2) kmer consistency, and 3) concordance to a benchmark, where discordances are interpreted as assembly errors.

Genome in a Bottle (GIAB) provides a benchmark for the human genome with estimated accuracy of 99.9999% (Q60). Concordance for human HiFi assemblies exceeds Q50, which provides excellent genomes for downstream analysis, but presents a challenge that any new benchmark must significantly exceed Q50 or the discordance will represent the error rate of the benchmark.

To establish benchmarks for *Oryza sativa* and *Drosophila melanogaster*, we collected draft references, Illumina short reads, and PacBio HiFi reads. By species, the benchmark was defined as regions of normal coverage that are not within 5 bp of a small variant or 50 bp of a structural variant. For both species, the benchmark regions span around 60% of the genome and HiFi assemblies achieve Q50 accuracy, which is notably more accurate than assemblies with other technologies and meets typical standards for a finished, reference-grade assembly.

Here we present a protocol to generate benchmarks for any sample that rival the GIAB benchmark in accuracy. These benchmarks allow the comparison and improvement of genome assemblies and highlight the superior accuracy of assemblies generated with PacBio HiFi reads.

PO0111: Methods: Sequencing

Next Generation Genotyping (NGG) for Population Scale Genomic Studies Using Riptide™ DNA Library Preparation

Keith Brown¹, Azeem Siddique¹, Gaia Suckow², Nils Homer³, Phil Ordhoukhanian², Steve Head², Joe Pickrell⁴ and Ryan Kim⁵, (1)iGenomX, Carlsbad, CA, (2)The Scripps Research Institute, La Jolla, CA, (3)Fulcrum Genomics, Somerville, MA, (4)Gencove, New York, NY, (5)MacroGen Corp., Rockville, MD

Over the past two decades, large population genotyping studies have delivered a better understanding of disease, an expanded pipeline of therapeutic targets and new diagnostic/predictive tests to improve health and wellness. These advancements are a result of innovation in genotyping technology. From realtime PCR and TaqMan assays through microarrays and sequencing, the drive for more information at a reduced cost has been consistent. Today's most comprehensive microarrays can provide over 5 million single nucleotide variants (SNVs) at a cost of approximately \$0.001 per genotype (\$450 / sample). We are presenting a Next-Generation Genotyping (NGG) approach and a novel DNA library prep product that enables the sequence-based genotyping of more than 37 million genetic markers at a cost of less than \$0.000002 / genotype (less than \$80/sample). In addition to the dramatic increase in marker density throughout the genome, NGG avoids the ascertainment bias associated with static content microarrays.

NGG is enabled by high-throughput sequencers with the main application criteria being low error rates and high read counts. As the widespread adoption of NGS platforms for applications like NIPT has shown, achieving high quality results at low cost is simply a matter of obtaining just enough evidence (defined as number of reads) to make high confidence calls on known targets. Sequencer reads are mapped to a reference genome, a haplotype is identified, and genotypes are assigned. Twenty plus years of sequencing and genotyping have made this approach

robust today. In this study, we evaluate the minimum performance criteria required for genotyping to high precision and sensitivity, using the iGenomX RipTide high throughput DNA library preparation product.

PE0112: Methods: Sequencing

Ultralow Input Single Tube Linked Read Technology Enables Short Read NGS Systems to Generate Highly Accurate and Economical Long-Read Information for Haplotype Phasing and De Novo Sequencing

Tom Chen¹, Long Pham¹, Gloria Mo¹, Donna Wu¹, Yu Xia¹, Peter Chang¹, Tan Phan², Huu Che², Son Pham², Yong Wang¹, Ming Lei¹ and Arthor Group, (1)Universal Sequencing Technology, Carlsbad, CA, (2)Bioturing Inc, San Diego, CA

Long read sequencers have been improved recently. However, they are still more expensive on the run cost and lower on the raw read accuracy than those from short read NGS system. In addition, microgram material input requirement for library preparation of these long-read sequencers significantly restrains their usage for many real-world samples. TELL-SeqTM linked read library technology overcomes all these challenges and enables highly accurate short NGS reads for long read applications economically. With less than nanogram ultralow sample input and highly scalable 3-hour library workflow, TELL-Seq extended long read capability into many previous unachievable applications using short read sequencing systems. With the single tube TELL-Seq method and 1 ng input DNA, we were able to *de novo* assemble *Arabidopsis* and *Drosophila* genomes. Other sizes of genomes ranging from megabases to gigabases were able to be *de novo* sequenced with 0.5ng to 5ng genomic DNA input. We also phased NA12878 and NA24385 human samples with N50 phased block size over 15 Mb long. Using phased reads, TELL-Seq technology significantly improved structural variation detection in these samples while kept both SNP variant calling sensitivity and precision at 99%. TELL-Seq technology has over 2.4 billion barcode diversity and is capable to assign one unique barcode sequence to only one long fragment and its associated sub-fragments for the highest barcoding specificity when necessary. With this highest level of barcoding resolution, we have demonstrated TELL-Seq advantage on metagenomic application and haplotype phasing of targeted contiguous gene loci using subnanogram input material.

PO0113: Methods: Sequencing

High MW DNA Extraction From Diverse Animal and Plant Species

Michelle Kim, Renee Fedak, Duncan Kilburn, Jeffrey Burke and **Kelvin Liu**, Circulomics Inc, Baltimore, MD

Long-read sequencing technologies from PacBio and Oxford Nanopore have become the standard methods for *de novo* sequencing due to the high contiguity of the resulting assemblies. 10X Genomics linked-read sequencing and Bionano Genomics optical mapping can be used to further refine assemblies. However, a significant bottleneck in these methods is the ability to obtain high purity, high molecular weight (HMW) DNA from challenging sample types.

We present new protocols for HMW DNA extraction from diverse animal and plant species using Nanobind technology. Nanobind magnetic disks are covered with a high density of micro- and nanostructured silica that protects DNA from damage to enable extraction of HMW (300+ kb) and UHMW (1+ Mb) DNA via a simple bind, wash and elute process.

First, new tissue protocols are presented that generate high purity, HMW DNA across mammalian, fish, avian, crustacean, and mollusk samples. Second, two plant protocols are compared, one based on nuclei isolation and a new protocol based on direct tissue lysis. Then, a new protocol for isolation of HMW DNA from insects is presented. Extraction results are shown from representative members of each sample category along with comparisons of fresh, frozen, and preserved sample types. Finally, sequencing results on PacBio Sequel I/II and Oxford Nanopore MinION/GridION are shown. Consistent read length N50 (20-60 kb) and data throughput are demonstrated across nearly all sample types. In aggregate, these methods give researchers a single toolkit that can generate reliable results across diverse samples.

PE0114: Methods: Sequencing

The Xdrop™ Technology: Targeted Enrichment of Single Molecules Resolves Complex Genomic Structures

Henrik Pfundheller, Samplix, Denmark

Long-read sequencing can resolve regions of the genome that are inaccessible to short reads, and therefore such technologies are ideal for genome-gap closure, solving structural rearrangements and sequencing through repetitive elements. Here we introduce the Xdrop technology: a novel microfluidic-based system that allows for targeted enrichment of long DNA molecules starting from only a few nanograms of genomic DNA. Xdrop is based on isolation of long DNA fragments in millions of droplets, where the droplets containing a target sequence of interest are fluorescently labeled and sorted using flow cytometry. The final product from the Xdrop procedure is an enriched population of long DNA molecules that can be investigated by sequencing.

To demonstrate the capability of Xdrop, we performed enrichment of the human papilloma virus (HPV) 18 integrated in the genome of human HeLa cells. The enriched DNA was sequenced both on long-read (PacBio and Oxford Nanopore) and short-read (Illumina) platforms. Analysis of the sequencing reads resolved three HPV18-chr8 integrations at base pair resolution, and the captured fragments extended up to 30 kb into the human genome at the integration sites. In summary, our results show that Xdrop is an efficient enrichment technology for studying complex regions of the genome where long-range information is required.

PO0115: Methods: Sequencing

Xdrop™ Strategies for Targeted Enrichment to Close Sequence Gaps in Poorly Assembled Genomes

Peter Mouritzen, Samplix, Herlev, Denmark

Large genome size and polyploidy of many eukaryotes have prevented whole genome sequencing and a thorough characterization of these genomes to the extent known from human and some primates. Apart from the bare size of some eukaryote genomes, regions with repeats and high GC content remain a challenge for the next generation sequencing technologies. The consequence is that many draft genomes include mis-assemblies and gaps, making it difficult to understand genetics and utilize genomic information e.g. in precision breeding even for some of the eukaryotes with moderate genome sizes. Long read sequencing technologies like Oxford Nanopore and PacBio are able to sequence through repeats and GC-rich regions but the costs of whole genome sequencing can be prohibitive.

Here we present Xdrop™, a new microfluidic droplet-based technology for targeted enrichment of genomic regions up to 100-150kb, compatible with downstream analysis by both long and short read sequencing technologies. The Xdrop™ enrichment is unaffected by repeats or high GC content and enables gaps-closing in unknown genomic regions. The only requirement for the enrichment is that a primer set can be designed to generate a short 150-180bp amplicon from a region neighboring the unknown region. The limited requirement for the known sequence also makes the approach ideal for resolving genomics landscapes with only partial sequence information, such as the one deriving from mRNA transcripts.

Here we describe the potential of the Xdrop™ enrichment technology to resolve sequence gaps, characterize CG-rich or repetitive regions, distinguish genes from pseudogenes, analyze paralogues, and phase alleles to significantly improve poorly assembled genomes.

PE0116: Methods: Sequencing

Nebnext® Ultra™ II FS DNA: A Robust, Enzyme-Based DNA Library Preparation Method Compatible with Plant Samples.

Karen R. McKay, Lynne M. Apone, Vaishnavi Panchapakesa, Pingfang Liu, Bradley W. Langhorst, Timur Shtatland, Christine J. Sumner, Fiona J. Stewart, Eileen T. Dimalanta and Theodore B. Davis, New England Biolabs, Ipswich, MA

Next generation sequencing (NGS) is currently an important tool used for assembly polishing, genotyping, and other applications relevant to plant biology. DNA fragmentation is one of the initial critical steps in the construction of quality libraries for NGS. However, current fragmentation methods create a bottleneck on library preparation throughput. To meet this challenge, we have developed a robust library construction method (NEBNext Ultra II FS) that integrates enzyme-based DNA fragmentation with end-repair and dA-tailing in a single step, followed by adaptor ligation in the same tube. This method eliminates the need for expensive equipment to fragment DNA; moreover, the optimized workflow reduces the numerous cleanup and liquid transfer steps, reducing time, cost and errors associated with library construction.

The robustness of the Ultra II FS DNA library preparation workflow was tested using genomic DNA from a variety of plant sources including the model organism *Arabidopsis thaliana*, *Oryza sativa*, and *Zea mays*. Libraries were prepared from a range of DNA input amounts to achieve different insert sizes with or without PCR amplification. All libraries were sequenced, reads aligned to the appropriate reference genome, and quality metrics generated using Picard tools. Compared with the traditional, mechanical shearing based library preparation method, Ultra II FS is significantly easier to automate, has higher library conversion rate and similar or superior sequencing quality. These advances will allow greater use and adoption of NGS technologies in plant research.

PO0117: Methods: Sequencing

The Past, Present, and Future of Arima Hi-C for Genome Assembly: A Focus on Sample Prep and Genome Scaffolding

Anthony Schmitt, Stephen Mac, Christa Caneda, Derek Reid, Xiang Zhou, Melissa Won and Siddarth Selvaraj, Arima Genomics, San Diego, CA

Generating a high-quality genome assembly is a critical starting point towards understanding the biology of an organism. A predominant approach to genome assembly involves scaffolding contigs into chromosomes, often leveraging chromosome-spanning linked-reads obtained from Hi-C. This value of Hi-C data has resulted in its broad utilization across assembly projects, including the Vertebrate Genome Project (VGP). Our objective is to provide superior Hi-C products, services, and support to serve the genome assembly community. Here, we demonstrate the present utility of high quality Arima-HiC data for chromosome-scale scaffolding of vertebrate genomes through our involvement as a VGP Phase I technology partner. We also report our ongoing analyses of sample collection and preservation methodologies for high quality Arima-HiC analysis, in collaboration with the VGP team. Lastly, we report significant new developments manifested in our next generation Arima-HiC 2.0 technology to meet the evolving needs of the plant and animal genome assembly community.

To best serve the genome assembly community with consistent high quality Hi-C data, we developed Arima-HiC technology. Our kits and services incorporate optimized protocols for various sample sources (plants, animals), types (tissue, blood, cells), preservation methods (snap-frozen, ethanol, RNAlater) and quantities. Through our ongoing partnership with the VGP as of July 2019, we have delivered 56 high quality Arima-HiC datasets from tissue and blood samples derived from 18 fish, 22 birds, 11 mammals, 4 reptiles, and 1 amphibian. Of these, 39 genomes have incorporated Arima-HiC data to produce high quality draft or complete annotated assemblies, all with >10Mb scaffold NG50s (AVG=73.9Mb). We have also systematically benchmarked sample collection temperatures, archival periods, and preservation reagents. These experiments have defined a set of compatible and optimal protocols for researchers to use in the field for sample collection and preservation.

We have also focused on meeting the unique requirements of forthcoming assembly projects by developing automated and ultra-low input workflows and a new Hi-C chemistry (termed Arima-HiC 2.0) with uniform per-base genome coverage designed to improve scaffolding and haplotype phasing and enable Hi-C data for use in base polishing analyses. Our new Arima-HiC 2.0 technology, in collaboration with select external partners, has been validated to produce an abundance of high-quality chromosome-spanning linked-reads in human, bird, mammals, and insects, with ongoing evaluations in plants. *In silico* analyses of genome coverage across 26 vertebrate, insect, plant, and parasite genomes indicate that at least 98.4% (STDEV=1.1%) of genomic sequence is accessible using our 2.0 chemistry. Deep sequencing analysis of Arima-HiC 2.0 data in human genomes reveals highly uniform coverage, with the per-base coverage being more uniform than depth-matched 10X Linked-Reads, and, equivalent SNV sensitivity when compared to shotgun Illumina sequencing.

In sum, Arima-HiC technology is a powerful and flexible tool for generating high quality Hi-C datasets and accurate chromosome-scale genome assemblies. The technology has been broadly validated, such as the extensive utilization of Arima-HiC technology to scaffold VGP genomes. We are continuously developing additional workflow and sample prep innovations to meet the growing needs of genome assembly projects.

PE0118: Methods: Sequencing

PalaeoChip: A Capture Enrichment Approach to Ancient Environmental DNA

Tyler J. Murchie^{1,2}, **Jennifer Klunk**³, Melanie Kuch^{1,2}, Ana T. Duggan^{1,2}, Marissa L. Ledger⁴, Kévin Roche^{5,6}, Emil Karpinski^{1,7}, Dirk Hackenberger^{1,8}, Tara Sadoway⁹, Ross MacPhee¹⁰, Duane Froese¹¹ and Hendrik Poinar¹, (1)McMaster Ancient DNA Centre, Hamilton, ON, Canada, (2)Department of Anthropology, McMaster University, Hamilton, ON, Canada, (3)Arbor Biosciences, Ann Arbor, MI, (4)University of Cambridge, Department of Archaeology, Cambridge, United Kingdom, (5)University of Bordeaux Montaigne, Pessac, France, (6)University of Bourgogne Franche-Comté, Besançon, France, (7)Department of Biology, McMaster University, Hamilton, ON, Canada, (8)Department of Biochemistry, McMaster University, Hamilton, ON, Canada, (9)University Health Network, Toronto, ON, Canada, (10)Division of Vertebrate Zoology/Mammalogy, American Museum of Natural History, New York, NY, (11)Earth and Atmospheric Sciences, University of Alberta, Edmonton, AB, Canada

Two major limitations in palaeoenvironmental metagenomics are DNA loss during extraction and the carryover of enzymatic inhibitors. PCR metabarcoding can overcome some degree of inhibition, but this technique can be vulnerable to differential amplification rates and a subsequent bias in taxonomic profiles, especially if there was substantial DNA loss with overly thorough inhibitor removal treatments. Alternatively, sedimentary ancient DNA (sedaDNA) extracted with techniques designed to maximize ancient DNA recovery and genomic coverage are prone to the carryover of enzymatic inhibitors, which can result in failed adapter ligation, impeding shotgun and targeted enrichment strategies.

Here, we report on a new sedimentary ancient DNA extraction protocol paired with targeted enrichment for reconstructing past environments. Our approach averages a 14.6-fold increase in on-target plant and animal DNA compared to a commercial soil extraction kit, and a 22.6-fold increase compared to a PCR metabarcoding approach. To illustrate the effectiveness of our PalaeoChip protocol, we present results of plant and animal eDNA capture enrichment from Yukon permafrost cores dating from the Last Glacial Maximum to early Holocene, along with new potential evidence for the late survival (ca. 9685 BP) of mammoth (*Mammuthus* sp.) and horse (*Equus* sp.) in the Klondike Region of Yukon, Canada. Our approach translates to a more diverse dataset with increased sequencing efficiency of ecologically informative sedaDNA.

PO0119: Methods: Sequencing

Next-Generation Target Capture and Direct Long Read Sequencing of Complex Regions

Johannes Dapprich, Generation Biotech / GenXPro, Princeton, New Jersey

The ability to selectively capture and sequence large, contiguous chromosomal segments based on very small target footprints is a significant advancement towards the comprehensive characterization of complex genomic regions, especially for highly repetitive or duplicated sequences that typically are very difficult to isolate and analyze unambiguously.

We show that DNA segments in excess of 20 kb in length captured via Region-Specific Extraction (RSE) can be processed and sequenced on Oxford Nanopore Technologies' MinION long-read platforms. We further demonstrate that with sufficient amounts of DNA this process can be carried out directly, requiring no amplification, thereby generating phased methylated base information over kb distances.

RSE can be performed either by hand or on automated DNA preparation devices and has proven to generate much longer sequencing templates and higher coverage when compared to other enrichment technologies. Due to the long read lengths isolated by RSE, the number of positions needed to target a particular region is 10-100-fold reduced

versus other capture methods. For instance, only 500 capture primers are required to reliably isolate the entire 4.8 Mb human major histocompatibility complex (MHC). One can create near-universal primer sets based on conserved reference points to successfully capture a highly polymorphic target region across many different DNA samples, irrespective of its complexity and diversity. Short (15-25) sequence-specific primers are hybridized to the target region after denaturing genomic DNA. An enzymatic extension with biotinylated nucleotides then generates 'handles' through which the underlying long DNA segments are securely captured. The DNA is then used for standard library preparation and sequencing, for instance on the MinION or any other NGS platform. Methylated bases can be directly detected without amplification if sufficient genomic DNA is used as input material.

Our study provides evidence that RSE coupled to long-read sequencing technologies allows to

- 1) completely and accurately characterize any genomic region, facilitating de-novo assembly and the identification of structural variants,
- 2) establish extended haplotypes through highly GC-rich or repetitive regions,
- 3) eliminate artefacts from unintended self-priming events during whole genome amplification,
- 4) generate orders of magnitude greater throughput than whole genome sequencing, and
- 5) directly detect base-modifications (methylation) without the need for a bisulfite conversion.

PE0120: Methods: Sequencing

Rapid Purification of Long-Read DNA for Nanopore Sequencing of Plant and Animal Samples

Shaun Veran, Zymo Research, Irvine, CA

Nanopore and PacBio Sequencing platforms can generate very long-read DNA sequences with incredible speed, flexibility, and resolution. However, extracting high molecular weight DNA suitable for these platforms has remained a time-consuming and tedious process that can take up to 2 hours. Here we present a rapid method to extract high molecular weight DNA up to 200 kb in just 40 minutes. The magnetic bead-based chemistry allows for an automatable, ultra-pure, high molecular weight DNA purification that is ideal for all third-generation sequencing platforms including Nanopore and PacBio. This is exemplified here with Nanopore sequencing of plant tissue, mouse tissue, and human blood samples.

PO0121: Methods: Sequencing

Balancing Mitochondrial and Genomics Sequencing Coverage in Targeted GBS Applications

Angela Burrell, Haktan Suren, Prasad Siddavatam and Rick Conrad, Thermo Fisher Scientific, Austin, TX

Mitochondrial SNPs are frequently used in targeted genotyping-by-sequencing (GBS) for the identification of individuals in forensic applications. However, having hundreds of copies of mitochondrial DNA (mtDNA) per cell as compared to only two copies of genomic DNA (gDNA) poses a challenge in next generation sequencing. The relatively high abundance of mtDNA can lead to over-representation during sequencing and poor coverage of gDNA targets. We have developed a strategy for combining mtDNA and gDNA targets in the same SNP panel while maintaining good coverage balance across all amplicons.

Two strategies were tested for optimizing co-detection of mtDNA and gDNA SNPs during sequencing reactions. We tested serially diluting either the mtDNA AgriSeq™ primer panel or the mtDNA amplicons between 1:4X and 1:200X relative to the gDNA. Libraries were then prepared using the standard AgriSeq™ HTS Library Kit and sequenced on the Ion GeneStudio S5 Sequencer using a 540 chip. Mean coverage was compared across all amplicons to find a dilution that resulted in balanced coverage between the mtDNA and gDNA targets. The best

results were seen when using a 1:8X dilution of the mtDNA AgriSeq primer panel. Diluting the primer panel instead of the resulting amplicons also had the advantage of allowing all amplification reactions to occur in a single well instead of separate reactions for mtDNA and gDNA amplification.

The optimized protocol was tested with a 119-SNP panel (14 mtDNA and 105 gDNA targets) against 102 canine samples in replicates. The mean call rate was >98% and the replicate genotype concordance was >99.8%. Coverage was very well balanced between mtDNA and gDNA targets with <0.2X difference in mean coverage between them. In conclusions, we have developed an optimized workflow to balance coverage between mitochondrial and genomic targets for optimal performance from both DNA sources.

For Research Use Only. Not for use in diagnostic procedures.

PE0122: Methods: Sequencing

A Flexible Automation Solution for Genotyping By Sequencing in Plant Breeding to Maximize Sample Throughput

Michelle S Swimley, Roy C. Willis, Angela Burrell and Rick Conrad, Thermo Fisher Scientific, Austin, TX

Marker assisted breeding using targeted genotyping by sequencing (GBS) is gaining traction as an effective tool for advanced breeding. We have developed and validated a 1,536-barcode set for multiplexed sequencing using AgriSeq™ targeted GBS technology. While 1,536 barcodes provide a tremendous potential sample throughput, the logistics of handling four 384-well plates of barcoded samples can be arduous and time consuming to perform manually. The use of a traditional liquid handler can reduce hands-on-time, however, the number of tips required for processing large numbers of samples can be negatively impactful, both economically and environmentally. To mitigate this impact, we have incorporated a MANTIS® liquid handler from Formulatrix® into the AgriSeq™ workflow. The Mantis is a positive air-displacement system that precludes the use of disposable tips and delivers the necessary volumes for repetitive dispersals of enzyme mixes, binding solutions, and washes; saving cases of pipet tips and reducing reagent ‘dead’ volume while easing operator fatigue and potential for technical errors. Capable of delivering a full 384-well plate’s worth of reagents in 5-10 minutes, this non-contact dispenser results in significant time savings as well. We have verified that performing the library preparation protocol with the MANTIS® provides results that are equivalent to or better than a purely manual workflow or one using a more traditional liquid handler while lessening our environmental impact through reduction of tips used. Sequencing 1,536 unique libraries on a single chip in less than 3 hours on the Ion GeneStudio S5 Prime system makes AgriSeq™ targeted GBS technology more efficient and affordable.

For Research Use Only. Not for use in diagnostic procedures.

PO0123: Methods: Sequencing

Neb Next Direct Genotyping Solution: Application of a Novel Method for Highly Efficient Targeted Genotyping across a Wide Range of Plant Species

Sarah K. Bowman¹, Amy B. Emerman¹, Kruti M. Patel¹, Brendan S. Desmond¹, Jonathon S. Dunn¹, Scott M. Adams¹, Susan E. Corbett¹, Charles D. Elfe¹, Bjoern Textor², Andrew Barry³, Theodore B Davis³, Felix Bemm⁴, Anja Hohmeyer⁵, Tereza Borba⁶, Evan Mauceli¹ and Cynthia L. Hendrickson⁷, (1)Directed Genomics, Ipswich, MA, (2)New England Biolabs GmbH, Frankfurt am Main, Germany, (3)New England Biolabs, Ipswich, MA, (4)KWS SAAT SE & Co. KGaA, Einbeck, Germany, (5)Trait Genetics, Stadt Seeland OT Gatersleben, Germany, (6)Embrapa Rice and Beans, Santo Antonio de Goias, Brazil, (7)Directed Genomics, Cambridge, MA

Targeted genotyping by next generation sequencing has significant advantages over traditional genotyping methods for marker assisted breeding, including the ability to interrogate thousands of markers at high resolution while discovering newly introduced SNPs that may contribute to phenotypic changes. Plant genotyping presents additional

challenges in target enrichment design, library preparation, and data analysis, arising from incomplete or absent reference genomes, repetitive sequences, and the presence of enzyme-inhibiting contaminants in DNA samples.

The NEBNext Direct Genotyping Solution is a novel method that combines highly multiplexed, capture-based enrichment with next generation sequencing (NGS) to deliver a one day, automatable genotyping workflow for high throughput breeding applications. Providing highly specific capture of marker ranges spanning 100 to 5,000 markers, pre-capture multiplexing of up to 96 samples combined with dual-indexed plus UMI sequencing allows for more than 3.8 million genotypes in a single Illumina® sequencing run.

Here, we demonstrate the successful application of the NEBNext Direct Genotyping Solution across plant species representing a range of genome sizes and ploidies to support high throughput breeding applications.

PE0124: Methods: Sequencing

MAUI-Seq: Metabarcoding Using Amplicons With Unique Molecular Identifiers to Improve Error Correction

Bryden Fields¹, Sara Moeskjær², Ville-Petri Friman¹, **Stig Uggerhøj Andersen²** and J. Peter W. Young¹, (1)University of York, York, United Kingdom, (2)Aarhus University, Aarhus, Denmark

Sequencing and PCR errors are a major challenge when characterising genetic diversity using high-throughput amplicon sequencing (HTAS). We have developed a multiplexed HTAS method, MAUI-seq, which uses unique molecular identifiers (UMIs) to improve error correction by exploiting variation among sequences associated with a single UMI. We show that two main advantages of this approach are efficient elimination of chimeric and other erroneous reads, outperforming DADA2 and UNOISE3, and the ability to confidently recognise genuine alleles that are present at low abundance. MAUI-seq provides sensitive and flexible profiling of diversity and is readily adaptable to most HTAS applications, including microbial 16S rRNA profiling and metabarcoding of environmental DNA.

PO0125: Methods: Sequencing

A Customizable Approach for Selective Removal of Abundant RNAs Enhances the Sensitivity of Transcript Detection Across Species

Deyra N. Rodriguez, Bradley W Langhorst, Kaylinnette Pinet, Keerthana Krishnan, Gautam Naishadham, Lynne M. Apone, Fiona J Stewart, Eileen T Dimalanta and Theodore B Davis, New England Biolabs, Ipswich, MA

The large dynamic range of transcript expression within total RNA presents a challenge to whole-transcriptome sequencing. Highly expressed transcripts with minimal biological interest can dominate readouts, masking detection of more informative lower abundance transcripts. Here, we present a customizable approach to enrich for RNAs of interest by eliminating unwanted RNAs. This method is based on hybridization of probes to the targeted RNA and subsequent enzymatic degradation of the selected RNAs. The probe sequences confer RNA removal specificity and can be designed to deplete unwanted RNA from any organism.

We developed a user-friendly web tool to enable custom depletion of any RNAs of interest. We used this web tool and depletion method to remove rRNA from total RNA of various species, including archaea *Thermococcus kodakarensis* and *Pyrococcus furiosus*, as well as the mosquito *Aedes aegypti*. Additionally, we used this approach to target highly abundant coding RNAs in human total RNA, and supplemented an existing anti-rRNA probe set for depletion of both rRNA and the selected coding RNAs.

Using strand-specific RNA sequencing we measured depletion efficiency and transcript expression. We achieved high depletion efficiency (up to 99%) for all targeted RNAs across species, while maintaining transcript abundance of non-targeted RNA. This translated into an enrichment of RNAs of interest and an increased depth of sequencing coverage.

The method described here is a simple and reliable solution to improve sensitivity in RNA-Seq studies. The design tool offers flexibility and control over the probe design process facilitating a customized and economical RNA sequencing experience.

PE0126: Methods: Sequencing

Selective Removal of Abundant RNAs Increases Transcriptome Profiling Sensitivity in Flowering Plants

Amanda Hoppers, Deyra N. Rodriguez, Bradley W Langhorst, Lynne M. Apone, Fiona J Stewart, Eileen T Dimalanta and Theodore B Davis, New England Biolabs, Ipswich, MA

RNA-seq is a widely used technology with a broad range of applications including differential expression analysis, alternative splice forms identification, etc., in both normal and disease contexts as well as in developmental studies. The technology has been pushed to extremes of very low and degraded sample inputs, but still battles with the challenge of having a large dynamic range of transcript expression. Highly expressed transcripts with minimal biological interest can dominate readouts, masking detection of more informative lower abundant transcripts. Here we present a method to enrich for RNAs of interest by eliminating abundant, typically unwanted RNAs.

This method, based on hybridization of probes to the target RNA and subsequent enzymatic degradation of the bound RNAs and the probes, has been optimized to remove abundant rRNAs from Mesangiospermae total RNA. We applied this method to deplete rRNAs from corn, rice, soybean and *Arabidopsis*. We measured sequencing metrics before and after depletion using strand-specific RNA sequencing. Across all samples we achieved high depletion efficiency while maintaining transcript abundance. The method works efficiently with a wide range of inputs from 10ng-1ug of total RNA, as well as degraded total RNA.

We conclude that the reduction of abundant transcripts for RNA-Seq studies significantly increases the ability to detect true biological variations that could not be detected in non-depleted samples. The method described here is a reliable and simple solution that greatly improves sensitivity in transcriptome RNA-Seq studies and is amenable to high throughput automation.

PO0127: Methods: Sequencing

3D RNA-Seq - A Powerful and Flexible Tool for Rapid and Accurate Differential Expression and Alternative Splicing Analysis of RNA-Seq Data for Biologists

Wenbin Guo¹, Nikoleta Tzioutziou¹, Gordon Stephen², Iain Milne², Cristiane Calixto¹, Robbie Waugh², John Brown¹ and **Runxuan Zhang**², (1)University of Dundee, Dundee, United Kingdom, (2)The James Hutton Institute, Dundee, United Kingdom

RNA-seq allows the quantification of expression of individual genes and transcripts and detection of alternative splicing (AS). Many RNA-seq differential analysis programs are unable to handle complex experimental designs (such as time-course data), require dedicated bioinformaticians and currently take months to obtain the results of the analysis. RNA-seq, despite its potential for accurate and rapid expression analysis at the transcript level, is often a source of frustration for biologists. We have developed “3D RNA-seq Shiny App”, which provides an easy and fast graphical user interface (GUI) for flexible and powerful Differential Expression (DE), Differential Alternative Splicing (DAS), Differential Transcript Usage (DTU) (3D) and isoform switch analysis of RNA-seq data in a matter of hours!! The program integrates the state-of-the-art, highly rated differential analysis tools (Limma) and adopts best practice for RNA-seq analysis. It can handle complex experimental designs, runs the analysis through a user-friendly GUI, visualizes the intermediate and final results through graphics and tables, and generates publication quality heat-maps, volcano plots, expression profiles and GO enrichment plots. The software is also available as a web service at <https://3drnaseq.hutton.ac.uk/>. The 3D RNA-seq tool and detailed step-by-step graphic manual makes it very easy to use for both experienced bioinformaticians and non-bioinformaticians. The tool allows lab scientists to “take back control” of the analysis of their RNA-seq data and generate results themselves.

The App takes transcript quantifications from Salmon or Kallisto, pre-processes the data (generating standardised read counts and reducing noise by removing low expressed transcripts and batch effects, and normalisation), sets up

statistical models with user-specified experimental factors and parameters, and runs and generates a customized and comprehensive analysis report with a click of the mouse. In a typical analysis, transcript quantification takes up to two days, and the analysis and report generation using 3D RNA-seq takes a few hours (3-Day RNA-seq!).

The App and pipeline have been successfully applied to RNA-seq data analyses in plants (e.g. Arabidopsis, potato, barley, etc.), human and animal research and enabled biologists to detect key and novel genes and transcripts under transcriptional and AS regulation. The App also allows biologists to re-analyse existing or publicly available RNA-seq data to give consistent and improved differential expression analysis and novel AS information. The acceleration of RNA-seq analysis with 3D RNA-seq App potentially reduces the time for a complete RNA-seq experiments to 3-4 months (RNA-seq data generation in 4-6 weeks, analysis in less than a week and interpretation in another 4-6 weeks). This allows multiple consecutive RNA-seq experiments to be conducted in a year. This speed of analysis will revolutionise what is achievable with RNA-seq technology. The 3D RNA-seq App won the Best Innovation Award in the University of Dundee, School of Life Sciences (2018-2019). Since the first release of the App and the pre-print manuscript in bioRxiv (<https://www.biorxiv.org/content/10.1101/656686v1>, May 31st, 2019), the App has ca. 1000 users with over 250 recurrent users from ca. 40 countries. The manuscript has over 13K pdf downloads.

PE0128: Methods: Sequencing

Evaluation of Transcript Assembly in Multiple Porcine Tissues Suggests Optimal Sequencing Depth for RNA-Seq Using Total RNA Library

Brittney N. Keel, William T. Oliver, John W. Keele and Amanda K. Lindholm-Perry, USDA, ARS, U.S. MEAT ANIMAL RESEARCH CENTER, CLAY CENTER, NE

RNA sequencing (RNA-Seq) libraries are prepared by either selecting polyadenylated (poly(A)) messenger RNAs or by depleting total RNA of highly abundant ribosomal RNAs. The latter facilitates novel transcript discovery by simultaneously characterizing both poly(A) and non-poly(A) RNA. Due to higher cost of preparing total RNA libraries, determining an optimal target sequencing depth would assist researchers in optimizing the cost-effectiveness of their experiments. The depth of sequencing needed for transcriptome profiling in total RNA-Seq was evaluated using a random sampling method. RNA-Seq was performed on 4 longissimus dorsi libraries, 4 liver libraries, and 8 hypothalamus libraries to produce base libraries of ~130 million (M) reads. Libraries were down-sampled to appropriate percentage of total reads to obtain approximately 5, 10, 20, 40, 60, 80, 100, and 120 M reads. Aligning reads to the genome and assembling transcripts resulted in 16,647 unique transcripts being identified in muscle, while 19,851 and 26,664 were identified in liver and brain, respectively. Multiple analyses highlighted the following: 1) sequencing depth below 10 M reads is insufficient for detecting transcript expression especially in medium and lowly expressed transcripts, 2) when sequencing depth is above 40 M reads relatively reliable measurement of expression is expected, and 3) sequencing deeper than 80 M reads does not have a significant effect on the number of transcripts identified. Thus, we propose that a depth of 80 M reads per library is sufficient to identify and quantify expression of transcripts across the genome. USDA is an equal opportunity provider and employer.

PO0129: Methods: Sequencing

A High-Quality PacBio Insect Genome from 5 Nanograms of Input DNA

Brendan Galvin¹, Douglas Shoue², Lei Zhu¹, Christine Lambert¹, Primo Baybayan¹, Sarah B Kingan¹, **Jonas Korlach**¹, Mary Ann McDowell² and Stephen Richards³, (1)Pacific Biosciences, Menlo Park, CA, (2)University of Notre Dame, Notre Dame, IN, (3)UC Davis, Davis, CA

High-quality insect genomes are essential resources to understand insect biology and to combat them as disease vectors and agricultural pests. It is desirable to sequence a single individual for a reference genome to avoid complications from multiple alleles during *de novo* assembly. However, the small body size of many insects poses a challenge for the use of long-read sequencing technologies which often have high DNA input requirements. The previously described PacBio Low DNA Input Protocol starts with ~100 ng of DNA and allows for high-quality assemblies of single mosquitoes among others, and represents a significant step in reducing such requirements. Here, we describe a new library protocol with a further 20-fold reduction in the DNA input quantity. Starting with just 5 ng of high molecular weight DNA, we describe the successful sequencing and *de novo* genome assembly of a

single male sandfly (*Phlebotomus papatasi*, the main vector of the Old World cutaneous leishmaniasis), using HiFi data generated from the PacBio Sequel II System and assembled with FALCON-Unzip. The assembly shows a high degree of completeness (97% complete BUSCO genes), contiguity (contig N50 of 1 Mb), and sequence accuracy (>98% of BUSCO genes without frameshift errors). We discuss the general utility of this workflow for the large number of small insects and other plant and animal species that can be best addressed using this method in the context of focused research studies or in conjunction with large-scale genome projects.

PE0130: Methods: Sequencing

Isolating High-Quality Ultra-High Molecular Weight (UHMW) Genomic DNA (gDNA) from 5-10 mg of Fresh or Freshly Frozen Animal Tissue

Yang Zhang, **Khoa Pham**, Henry B. Sadowski, Hannah Way and Mark Oldakowski, Bionano Genomics, San Diego, CA

Optical mapping of genomic DNA on the Bionano Genomics Saphyr® Whole Genome Imaging system for genome assembly or structural variation detection relies on starting with UHMW gDNA.

Here we present a streamlined Bionano Prep SP Animal Tissue DNA Isolation Protocol to isolate ultra-high molecular weight (UHMW) gDNA from solid tissue. It starts with a simplified front end tissue homogenization step for both soft and fibrous tissue types. The protocol then utilizes Bionano Prep SP where solution-based lysis is coupled with a purification step leveraging a novel process to bind, wash and elute UHMW genomic DNA. This entire protocol can be conducted in approximately six hours on a batch of six samples. The eluted material is ready to use by Day 2 and contains high quality DNA that is compatible with the Direct Label and Stain (DLS) Protocol. The resulting metrics of this labeled DNA on a Bionano Saphyr Chip® consistently result in high output runs (1.3 Tbp per sample) of high quality long molecules (>250kbp N50s) sufficient to be used in genome finishing, structural variant comparative analysis and QTL studies.

Using this method, we have successfully isolated UHMW DNA from 5-10 mg of fresh or fresh frozen animal and human tissues.

The method is very attractive compared to other protocols because: 1) it is amenable to automation, is less cost prohibitive, and provides a time saving solution for sample prep; and 2) uses very small amount of tissue - possible applications for rare samples and human tissue biopsy tests.

PO0131: Methods: Transformation

2020 World Congress on *In Vitro* Biology

Ashok Shrawat, Bayer, St. Louis, MO

This year's international Congress (June 6-10, 2020 at the Town and Country San Diego), held once every 4 years, is a phenomenal chance to explore the latest research and core principles of *in vitro* plant and animal work on a global scale through daily plenary sessions on *Emerging In Vitro Technologies; Bioethics and Public Policy Benefits in Genome Editing; Perspectives in Cannabis and Cannabinoids* and *Frontiers in Single Cell Technologies* in addition to Keynote on *Emergence of Spontaneous Oscillatory Networks from Human Brain Organoids* by Dr. Alysson R. Muotri from UC, San Diego. The program includes topics such as: *Application of Omic's Technology; Beyond KOs: Emerging Genome Editing Technologies; Exploring Microbiomes: Application to Humans and Agriculture; Organoid Models: Windows into Human Disorders; Plant Memory: The Importance of Assessing Culture Carry Over Effects During Micropropagation Protocol Development; and Plan(t)s for the Future Planet;* and more. In addition to hosting the upcoming 2020 SIVB Congress which provides scientists a venue to keep current in the fields of plant and animal biotechnology, SIVB also supports several additional activities each year that influence those in academia and industry. The society provides information on new technologies to policy makers and government agencies that bases its core in science. SIVB's membership actively educates the next generation of graduate students and postdoctoral researchers, helping them develop and refine skills required for quality jobs in both industry and academia.

PE0132: Methods: Transformation

Optimization of *in Planta* Transformation Methods in Rice (*Oryza sativa*)

Tia K. Dunbar, Nikolaos Tsakirpaloglou, Nancy Wahl, Endang M. Septiningsih and Michael J. Thomson, Texas A&M University, College Station, TX

Gene editing based on CRISPR/Cas technology offers the potential to modify crops in a timely manner to meet future global needs. However, most gene editing methods for crop improvement require time and labor-intensive *in vitro* tissue culture techniques, such as protoplast regeneration, callus induction, and root and shoot induction. Current transformation and regeneration techniques are also genotype-specific and not yet optimized for many minor crop species. Bypassing *in vitro* regeneration processes would facilitate gene editing and expand the use of this technology to a wide range of plant species. Our research investigates *in planta* transformation methods that can alter germline cells, thus eliminating the challenge of tailoring regeneration protocols specifically to each crop. Nanotechnologies present a promising approach: recent studies have demonstrated that carbon nanotubes (CNTs) loaded with DNA are able to penetrate the cell wall of fully intact plant cells. Passive diffusion of CNTs across the cell wall and nuclear membrane can facilitate transient expression of foreign genes in plant cells. CNTs delivering plasmids encoding Cas9 and guide RNAs into mature embryos can create genetic edits that, once the seed has matured into a full plant, can be inherited in the germline. We are currently testing these approaches to alter the phenotype of rice seedlings to visually test for successful plasmid delivery and gene edits. Streamlined *in planta* CRISPR reagent delivery and transformation will ultimately increase the capacity of gene-editing technologies as these CNT delivery methods can be applied to a diverse array of crops.

PO0133: Methods: Transformation

A Modified Protocol Improves Efficiency of *Agrobacterium*-Mediated Transformation in Cucumber (*Cucumis sativus* L.)

Hanqiang LIU, University of Wisconsin-Madison, MADISON, WI, Jianyu Zhao, Department of Horticulture, Madison, WI and Yiqun Weng, USDA ARS/ University of Wisconsin, Madison, WI

Genetic transformation is an important technique for gene function study and crop improvement, but in cucumber, its routine use remains challenging due to low transformation efficiency and poor reproducibility of published protocols. In this study, we revisited early protocols and tried to optimize technical details in the *Agrobacterium*-based transformation in cucumber. We used the North American slicing cucumber genotype Poinsett 76 as the test material and the *littleleaf (ll)* gene as the target. We found that the serrated and crescent wounds at the proximal part of the half cotyledons are preferred explants to induce shoots. The highest percentage of shoot regeneration was obtained when the explants were inoculated in the *Agrobacterium* (strain AGL1) suspension at OD₆₀₀ of 0.7-0.8 for 12 min facilitated with vacuum infiltration. Addition of filter papers on the co-culture medium seem to decrease *Agrobacterium* contamination significantly during shoot initiation. After co-culture for 2 days, transfer only yellowish explants to selection media containing 0.5mg/L 6-BA, 0.5mg/L ABA, 100mg/L kanamycin and 200mg/L timentin. Germinating of naked seeds at 28°C for 3d could significantly increase the portion of yellowish explants after co-culturing with *Agrobacterium*. No-chimeric shoots were sub-cultured in fresh selection medium for at least one more time. Only the seedling with green leaves grown from antibiotic selection were transferred to 1/2 MS rooting medium supplemented with 50mg/L kanamycin and 200mg/L timentin. Rooted plantlets with ~ three roots of ~2cm in length were transplanted to small pots to domesticate for 3d (covered with a plastic film to maintain moisture). The seedlings were then moved to the greenhouse for PCR verification. Only PCR positive plants were kept for phenotyping (T0) and production of progeny. Using this protocol, the average transformation efficiency for producing transgenic cucumber plants is ~1% (that is, 1 true transgenic plants from 100 seeds).

PE0134: Methods: Transformation

Establishment of *Agrobacterium* Mediated Transformation Protocol of *Chlorophytum borivilianum*: A Plant of High Medicinal Value

Nishant Kaushal, Anshu Alok and Kashmir Singh, Panjab University, Chandigarh, India

Chlorophytum borivilianum (Liliaceae family), is among the few rare medicinal plants witnessing steady growth in pharmaceutical and nutraceutical products. Due to high demand plant is over exploited and *in-vitro* regeneration

and multiplication methods for *C. borivilianum* have been established using different explants to meet the demand and supply. However, being a monocot, the transformation methods has not been developed properly for this plant.

Saponins are the principal active components of this plant and *Squalene epoxidase* (SE) is one of the important enzyme of the saponins biosynthetic pathway, so it has been cloned in pRI-101AN vector and then into different *Agrobacterium* strains to check the transformation efficiency. Further, this vector construct has been transformed into *Chlorophytum* and *Arabidopsis*. *Agrobacterium* harboring pCambia1305 plasmid carrying β -glucuronidase as a reporter gene and Hygromycin phosphotransferase as plant selectable marker genes was used for transformation of *C. borivilianum*. Up to 15% transient transformation was found with callus and shoots/leaves co-cultivated with *Agrobacterium* strains RT1022, K599, SA079 and AGL1301 for 2 days at 24 °C on M.S. media. Cefatoxime (200 mg/L) eliminated the bacterial growth in selection media. Transformation was confirmed by GUS histochemical assay and further by polymerase chain reaction (PCR) using *uidA* and *hpt* gene specific primers. Hygromycin lethality test showed concentration of 25 mg/L were sufficient to inhibit the growth of callus and multiple shoot buds of non transgenics in selection media. Also, In this study different growth regulators were considered for regeneration of multiple shoots and rapid callus growth. Explants (stem discs and inter nodal tissues) and callus were cultured on different growth regulators on Murashige and Skoog's medium containing growth regulators at different concentrations (BAP, 2,4-D, NAA and IAA) and observed after 60 days. Parameters such as *Agrobacterium* strains, O.D.₆₀₀ of bacterial culture, co-cultivation duration, acetosyringone concentration and explants type were optimized for transient expression of the reporter gene.

The study will be useful in improvement of *C. borivilianum* by genetic engineering for the production of large number of plants and to improve the saponins content in the plant by over expressing the gene *squalene epoxidase* using this transformation protocol.

PO0135: Methods: Transformation

Chloroplastic Transformation of Potato Cultivars to Encode Resistance Against Colorado Potato Beetle and Potato Tuber Moth

Allah Bakhsh¹, Md. Jakir Hossain¹, Faiz Ahmad², Emre Aksoy¹, Neslihan Zahide Öztürk Gökçe¹, Ghulam Mustafa² and Muhammad Sarwar Khan², (1)Nigde Omer Halisdemir University, Turkey, (2)University of Agriculture Faisalabad, Pakistan

Recently organelle transformation has become popular technology due to various advantages like high protein levels, the feasibility of expressing multiple proteins from polycistronic mRNAs, transgene stacking in operons, lack of epigenetic interference allowing stable transgene expression and gene containment through maternal inheritance of plastome. The present study is being carried out to transform potato chloroplasts with a hybrid insecticidal gene (*SN-19*) and *cry3A* gene under the control of psbA promoter in the pyramided form to encode resistance against notorious insect pests of potatoes ie. Colorado potato beetle and Potato tuber moth. Plastome sequences comprising mainly of trnA and trnI were used as flanking sequences of chloroplast transformation vector PCTV 9110. Further, this vector contains FLARE-S as selectable marker under the control of prn promoter, derived by the fusion of aadA and gfp genes to provide dual selection on plant regeneration medium. The different explants (leaf and internode) have been bombarded with constructed plasmids using particle bombardment method. Transplastomic lines have been regenerated using optimized in vitro culture protocol at our laboratory. The preliminary results of regeneration and foreign gene integration in transplastomic plants will be presented at meeting.

PE0136: Methods: Bioinformatics

AgBioData Consortium: Genomics, Genetics and Breeding (GGB) Databases Working Together

Jacqueline D. Campbell, Washington State University, Ames, IA, Ethalinda Cannon, USDA-ARS Corn Insects and Crop Genetics Research, Ames, IA, Lisa Harper, USDA ARS, Albany, CA, Dorrie Main, Department of Horticulture, Washington State University, Pullman, WA, Sook Jung, Washington State University, Pullman, WA, Monica Poelchau, USDA/Agricultural Research Service/National Agricultural Library, Beltsville, MD and AgBioData Consortium

The AgBioData Consortium of agriculture-related databases (<https://www.agbiodata.org>; Harper et al. 2018) was founded to ensure standards and best practices for acquisition, integration, display and retrieval of genomic, genetic and breeding data. Collectively, the AgBioData member databases served 27 million pages and 950,000 users in 2017, and between 2012-2017, they were cited in over 24,000 publications. The databases cover an extensive range of crops, livestock and model organisms, including arabidopsis, corn, fruit and nuts, legumes, vegetables, wheat, insects, cattle, chicken, fish, horse, pig, and sheep. To reach the goal of making every piece of biological data FAIR (Findable, Accessible, Interoperable, Reusable), both databases and researchers need to continue to work together to adopt good data management practices, a common set of metadata recommendations: associate data with ontologies; simplify data sharing; share data curation practices; and provide solutions for long-term funding for all genomic, genetic and breeding databases. AgBioData is a model for how databases can be a collective voice to lead efforts for better data management and data sharing.

PO0137: Methods: Bioinformatics

Using Genome Browsers Constructed by G-OnRamp to Provide Students with a Course-Based Undergraduate Research Experience in Genome Annotation

Wilson Leung¹, Luke Sargent², Yating Liu¹, Nathan T. Mortimer³, David Lopatto⁴, Jeremy Goecks² and Sarah C. R. Elgin¹, (1)Washington University in St. Louis, St. Louis, MO, (2)Oregon Health & Science University, Portland, OR, (3)Illinois State University, Normal, IL, (4)Grinnell College, Grinnell, IA

Course-based Undergraduate Research Experiences (CUREs) based on genome annotation are beneficial to researchers, educators, and students alike. They provide researchers with high quality gene models and provide educators with an effective way to teach students about eukaryotic genes/genomes. Genome browsers provide visualizations that facilitate the synthesis of multiple types of experimental and computational evidence for constructing gene models. To reduce the technical expertise required to construct genome browsers, the Genomics Education Partnership (GEP) and the Galaxy Project (<https://galaxyproject.org>) have developed G-OnRamp (<http://g-onramp.org>), a web-based platform for constructing UCSC Assembly Hubs and JBrowse genome browsers with evidence tracks for sequence alignments, gene predictions, RNA-Seq data, and repeats identification. G-OnRamp also provides tools to create and manage Apollo instances for collaborative genome annotations. G-OnRamp has been used to create genome browsers for >20 species (<http://g-onramp.org/genome-browsers>), including those for a CURE that examined lipid synthesis pathway genes in four parasitoid wasp species. This CURE engaged more than 200 students from 15 diverse institutions. Results from an anonymous survey of G-OnRamp users showed that most respondents find G-OnRamp useful in their research and their teaching; some plan to use it to develop new CUREs. Version 1.1 of G-OnRamp added the capability to incorporate extrinsic evidence into the Augustus gene predictions, and improved compatibility with new versions of Apollo, JBrowse, and Galaxy. G-OnRamp can be deployed locally via a virtual appliance or on the Cloud (Amazon EC2) via CloudLaunch (<http://g-onramp.org/deployments>). Faculty interested in developing a CURE using G-OnRamp can contact us at http://gep.wustl.edu/contact_us.

PE0138: Methods: Bioinformatics

The Genomics Education Partnership: Introducing Undergraduates to Research by Engaging Them in Genome Annotation

Sarah C. R. Elgin¹, Wilson Leung¹, Anne Rosenwald², David Lopatto³, Catherine Reinke⁴, Laura K. Reed⁵ and the Genomics Education Partnership, (1)Washington University in St. Louis, St. Louis, MO, (2)Georgetown University, Washington, DC, (3)Grinnell College, Grinnell, IA, (4)Linfield College, McMinnville, OR, (5)The University of Alabama, Tuscaloosa, AL

Since 2006, the Genomics Education Partnership (GEP; <http://gep.wustl.edu>) has helped faculty bring genomics research experiences into the undergraduate curriculum, introducing thousands of students to eukaryotic gene annotation, comparative genomics, and the evolution of contrasting genome domains. Our 100+ participating institutions include community colleges, primarily undergraduate institutions, and research-intensive PhD-granting institutions. Genomics research requires relatively little infrastructure and can utilize publicly accessible data, keeping costs low. With no safety issues, it lends itself to peer instruction, multiplying faculty effort. Student mistakes are inexpensive in both time and dollars, allowing students to fail in their original analysis, explore,

reevaluate, adjust, recover, and succeed. This iterative “formative frustration” process allows students to gain deeper insights into their research problem. Our research format encourages the use of active learning strategies and promotes dialogue about science, both of which contribute to positive student outcomes. Students are generally successful in generating defensible gene models, leading to co-authorship on scientific publications, but those with positive attitudes toward science benefit the most. To support additional collaborative annotation projects, GEP has partnered with Galaxy to develop G-OnRamp, a web-based platform for constructing genome browsers from new genome assemblies; we are looking for new “science partners” who have interesting projects for collaboration. We provide fully-supported training for new GEP faculty members through online mentoring and in-person workshops. The next in-person workshop is June 10-13, 2020, at Washington University in St. Louis. Those interested in joining the GEP can contact us at http://gep.wustl.edu/contact_us. Supported by NSF IUSE-1915544 and NIH IPERT-1R25GM130517-01 to LKR.

PO0139: Methods: Bioinformatics

Bioinformatics Workbook: Case Study

Siva Chudalayandi, Maryam Sayadi, Arun S. Seetharam and Andrew J. Severin, Iowa State University, Ames, IA

Bioinformatics and biological data analysis are an important aspect of biological research. Researchers in the genomic era ask larger questions that routinely require dealing with large amounts of data that require familiarity with medium to large scale computing. Bioinformatics can deal with complex and simple datasets. A good chunk of the algorithms in bioinformatics are devoted to aligning DNA and protein sequence data. For someone getting started in Bioinformatics, it is best learned through solving real world problems. With this in mind, we created a github based bioinformatics workbook (<https://bioinformaticsworkbook.org/>), that can serve as an online knowledge repository for scientists. We have covered bioinformatics analysis workflow for a variety of problems, like sequence alignment, genome assembly, comparative genomics, data visualization etc. We work through a specific problem, e.g., RNAseq, from raw data (downloaded from NCBI) through to the processed results. Additionally, for the beginner we have also included introductory lessons in the Unix command line and tips and tricks for common problems. Because novel techniques of analysis are developed frequently, we envision our workbook as an open source living book and invite peers to contribute to this venture.

PE0140: Methods: Bioinformatics

Livia: Evolving Pictures to Teach Evolution

Jason G Wallace, University of Georgia, Athens, GA

Evolution is the core of modern biology, but some concepts can be difficult for students to understand, especially if they come from a non-biological background. To help students understand how selection can turn a random process (mutation) into highly ordered, complex systems, I present Livia, an interactive R Shiny app where users can evolve a population of digital images to match a chosen target. Students can experiment with parameters such as mutation size, selection ratio, and trait complexity (grayscale/color) to see how each of these affects the rate and success of selection. In addition to basic evolutionary principles, this program can also be used to demonstrate more advanced principles such as mutation-selection balance and Fisher’s geometric model. Educators are encouraged to test this tool in the classroom, and to provide feedback on ways to make it more usable and accessible for students.

PO0141: Methods: Bioinformatics

The Galaxy Training Network: A Community Based Training Resource

Mo Heydarian, Johns Hopkins University, Baltimore, MD and The Galaxy Training Network

The Galaxy Training Network (GTN, <https://training.galaxyproject.org>) is a community based effort to provide training materials for commonly used bioinformatic pipelines using the Galaxy Workbench. The GTN has developed training resources for common applications in transcriptomics, epigenetics, variant analysis, genome assembly and annotation, and proteomics. The GTN tutorials have recently extended into additional domains including single cell transcriptomics, machine learning, computational chemistry, ecology, and even Galaxy administration. All GTN tutorials provide scientific context for the analysis application, small sample data for

demonstration, and detailed step-by-step instructions to complete a given analysis. Many of the GTN tutorials are accompanied by Galaxy Workflows and slides for research usage or teaching, respectively. The GTN training materials are hosted on Github (<https://github.com/galaxyproject/training-material>), where they are openly accessible and open to contributions.

PE0142: Methods: Bioinformatics

Galaxy: A Global Platform & Community for Life Science Data Integration & Analysis

Dave Clements, Johns Hopkins University, Eugene, OR and the Galaxy Community

Galaxy is a free and open source web-based data integration and analysis platform for life science research (galaxyproject.org). It is deployed in hundreds of organizations around the world and is supported by a global community of researchers, trainers, software engineers, and tool developers. Galaxy makes sophisticated bioinformatic analysis accessible and reproducible to researchers without requiring them to learn command line interfaces or computer systems administration. This poster will introduce several components of the Galaxy ecosystem and community, including

- The Galaxy graphical user interface, making tools and workflow construction accessible to bench researchers
- Galaxy Toolshed, containing over 7000 tools that have been ported to Galaxy
- Galaxy Platform Directory, listing over 150 platforms for using and running Galaxy
- Galaxy help resources include the online Galaxy Help forum, Gitter channels, and Galaxy Search
- Galaxy Events calendar, news feed, and regional and domain based communities
- Support tools and resources for deploying your own Galaxy instance.

PO0143: Methods: Bioinformatics

Quick and Easy Genome Annotation Editing with Apollo

Nathan A Dunn¹, Colin Diesh², Nomi Harris¹, Robert Buels³ and Ian Holmes³, (1)Lawrence Berkeley National Laboratory, Berkeley, CA, (2)University of California Berkeley, Bioengineering, Berkeley, CA, (3)University of California, Berkeley, Berkeley, CA

Precise descriptions of annotated genomes are vital for modeling the biological function of genomic elements. The ability of a researcher to visually identify and review diverse sets of information such as genomic and transcriptome alignments, predictive models based on sequence profiles, and predicted regulatory elements and repeat regions are essential for the iterative improvement of the modeling of genomic elements. During analysis, researchers also do functional enrichment analysis (such as GO), and need to update functional annotations. Furthermore, as projects increasingly include annotations of a growing number of organisms as well as geographically dispersed researchers, the ability to quickly integrate multiple genomes, sources of evidence, annotations and researchers is essential. The Apollo genome annotation editor fills these needs by providing a graphical platform for researchers to collaboratively review and revise the predicted features on a genome in real-time, similar to Google Docs. Refinement of genome annotations is made efficient through several features including drag-and-drop editing, a large suite of automated structural edit operations, the ability to pre-define curator comments and annotation status to maintain consistency, attribution of annotation authors, and a visual history of revertible edits.

Here, we describe recent improvements that increase the efficient refinement of genome annotations. The first is the automated processing of genomic evidence for annotation, reducing the need for command-line processing of genome annotation evidence. Creating annotation projects can be done by simply uploading the genome's FASTA file. Similarly, genomic annotation evidence can be provided in most cases by uploading GFF3, VCF, BigWig, and BAM files directly in most cases. The second is the ability to associate GO annotations to genome annotations and export in formats such as GPAD or GPI. The third is the ability to predict the effect of individual variants to aid in the annotation of variants. Finally, we demonstrate numerous UI improvements to make annotation editing faster and easier as well as the simplicity in launch Apollo from a simple Docker command or for preconfigured Community AMI on Amazon cloud instances.

In addition to the simplified installation process, Apollo provides extensive web-services that allow it to be integrated with other web-based environments. Apollo and its associated libraries allow numerous customizations, both within Apollo itself and via JBrowse, the genomic browser Apollo is built upon (<http://jbrowse.org>), which has a large library of plugins (<https://gmod.github.io/jbrowse-registry/>).

Apollo is used in hundreds of genome annotation projects around the world, ranging from the annotation of a single species to lineage-specific efforts supporting the annotation of dozens of genomes.

Source: <https://github.com/GMOD/Apollo/>

Documentation: <http://genomearchitect.readthedocs.io/en/latest/>

License: Open Source - 3-clause BSD License

PE0144: Methods: Bioinformatics

UCSC Genome Browser API

Brian Lee, Univ. Calif. Santa Cruz, Santa Cruz, CA

The UCSC Genome Browser has a new Application Programming Interface (API) with JavaScript Object Notation (JSON) output. The API provides direct access to different data including annotations and sequence data for both native Genome Browser assemblies as well as novel organisms that can be visualized within user-generated assembly hubs and accessed by the API. Some functions possible with the API include listing all available UCSC Genome Browser assemblies, listing chromosomes contained in a UCSC Genome Browser assembly or novel user-generated assembly hub, listing all data tracks in a specific assembly, extracting any track data from native Genome Browser assemblies or novel user-generated assembly hubs. A complete list of what data can be accessed, as well as examples of how to access the data, is available on the API help page:

<http://genome.ucsc.edu/goldenPath/help/api.html>

PO0145: Methods: Bioinformatics

MakeHub: Creating Individual Assembly Hubs for Display with the UCSC Genome Browser

Katharina J. Hoff, Institute for Mathematics and Computer Science, University of Greifswald, Greifswald, Germany; Center for Functional Genomics of Microbes, University of Greifswald, Greifswald, Germany

The UCSC Genome Browser [2] is one of the most powerful and most convenient tools for visualization of genomes with their annotation. Track data hubs [1] allow the display of externally hosted genomic data via publicly available UCSC Genome Browser instances. Creating your own assembly hub for a novel genome, however, is often a tedious task that involves many steps that are in part difficult for scientists without programming background.

MakeHub [3] has the goal to enable scientists to quickly and automatically generate assembly hubs of novel genomes, their annotation and informative RNA-Seq read alignments. Producing a complete assembly hub is a one-step process with MakeHub. Implemented in Python, MakeHub utilizes tools provided by the UCSC Genome Browser group, SAMtools [4], and components of the gene prediction tool AUGUSTUS [5]. MakeHub is integrated in the BRAKER [6,7] pipeline for fully automated and unsupervised RNA-Seq based structural genome annotation. It is further compatible with the outputs of MAKER [8], GlimmerHMM[9], SNAP [10] and GeMoMa [11].

MakeHub is freely available at <https://github.com/Gaius-Augustus/MakeHub>.

[1] Raney BJ, Dreszer TR, Barber GP, Clawson H, Fujita PA, Wang T, Nguyen N, Paten B, Zweig AS, Karolchik D, Kent WJ. 2014. "Track Data Hubs." *Bioinformatics* 1;30(7):1003-5

- [2] Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler D. 2002. "UCSC Genome Browser." *Genome Res.* 12(6):996-1006
- [3] Hoff KJ. 2019. MakeHub: Fully automated generation of UCSC Genome Browser Assembly Hubs. *Genomics, Proteomics and Bioinformatics*, in press; preprint at <https://www.biorxiv.org/content/10.1101/550145v2>
- [4] Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 2009. "The sequence alignment/map format and SAMtools." *Bioinformatics* 26(16):2078-2079
- [5] Stanke M, Diekhans M, Baertsch R, Haussler D. 2008. "Using native and syntenically mapped cDNA alignments to improve de novo gene finding." *Bioinformatics* 24(5):637-644
- [6] Hoff KJ, Lange S, Lomsadze A, Borodovsky M, Stanke M. 2015. "BRAKER1: unsupervised RNA-Seq-based genome annotation with GeneMark-ET and AUGUSTUS." *Bioinformatics* 32(5), 767-769
- [7] Hoff KJ, Lomsadze A, Borodovsky M, Stanke M. 2019. Whole-Genome Annotation with BRAKER. *Methods Mol Biol* 1962:65-95
- [8] Holt C, Yandell M. 2011. "MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects." *BMC Bioinformatics* 12(1), 491
- [9] Majoros WH, Salzberg SL. 2004. "TigrScan and GlimmerHMM: two open source ab initio eukaryotic gene-finders" *Bioinformatics* 20(16), 2878-2879
- [10] Korf I. 2004. "Gene finding in novel genomes" *BMC Bioinformatics* 5, 59
- [11] Keilwagen J, Hartung F, Paulini M, Twardziok SO, Grau J. 2018. "Combining RNA-seq data and homology-based gene prediction for plants, animals and fungi." *BMC Bioinformatics* 19(1), 189.

PE0146: Methods: Bioinformatics

NCBI Blast: Enhanced Web Usability Through New Result Page, and Effective Genomic Data Access

Tao Tao, NIH/NLM/NCBI, Bethesda, MD; National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD

NCBI provides BLAST services through its BLAST homepage (<https://blast.ncbi.nlm.nih.gov>) where the public can search against the nucleotide database (NT), the non-redundant protein database (NR), a set of NCBI Reference Sequence Project (RefSeq) genomic sequences, and their annotated RNA or protein sequences.

In this presentation, we will describe new features contained in the recently updated BLAST result pages that significantly enhance the page's usability. We describe the available RefSeq genomic databases and their relationship with genome records in NCBI's assembly database, which will enable more rational selection of databases and their taxonomic subset during a BLAST search. We will also cover alternative ways to access genomic sequences of interest, through assembly records on the web or FTP, for use in local standalone BLAST.

PO0147: Methods: Bioinformatics

Multi-Omics Approach to Pathway Analysis in Plant Genomes using Transcriptomics and Genome-Scale Metabolic Modeling in KBase.

Sunita Kumari¹, Vivek Kumar¹, Doreen Ware¹, Samuel M. D. Seaver², Christopher Henry², Robert W. Cottingham³ and Adam Arkin⁴, (1)Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, (2)Argonne

National Laboratory, Argonne, IL, (3)Oak Ridge National Laboratory, Oak Ridge, TN, (4)Lawrence Berkeley National Laboratory, Berkeley, CA

The Department of Energy Systems Biology Knowledgebase (KBase; <http://kbase.us>) is an open, web-accessible computational environment for systems biology research focused on microbes, fungal, plants and their communities. It provides a range of integrated biological data types and associated analysis tools (Apps) that include gene expression analysis and metabolic modeling. The user-friendly KBase Narrative Interface offers researchers and bioinformaticians a range of analysis tools and data resources that accelerate the pace of functional genomics research by allowing large-scale sample processing, expression-level quantification and integration of gene expression profiles with downstream functional analysis including ontology enrichment, and compare flux with expression

KBase currently has 124 plant genomes from the JGI Phytozome V13 database. KBase has several data resources that originated from the PlantSEED project. Plant-specific compounds and reactions, collected from public sources such as KEGG, MetaCyc, and AraCyc, have been integrated into PlantSEED and made available in KBase, where they can be used for plant metabolic modeling.

KBase is also actively engaged with the external community to help us improve the available tools and workflows for functional genomics especially support for gene expression, regulation, and epigenetics in plant science. These capabilities are directly relevant to important DOE research targets such as optimizing biomass production in biofuel feedstocks.

KBase is funded by the Genomic Science program within the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research under award numbers *DE-AC02-05CH11231*, *DE-AC02-06CH11357*, *DE-AC05-00OR22725*, and *DE-AC02-98CH10886*.

PE0148: Methods: Bioinformatics

TASUKE+: A Web-Based Platform for Large-Scale Resequencing Data and Exploring GWAS Results

Masahiko Kumagai, National Agriculture and Food Research Organization, Tsukuba, Japan

Motivation: Recent revolutionary advancements in sequencing technologies have made it possible to obtain mass quantities of genome-scale sequence data in a cost-effective manner and have drastically altered molecular biology studies. In particular, the genome-wide association study (GWAS) has become increasingly important. Hence, there is an urgent need to develop a browser that enables efficient retrieval and public dissemination of such large-scale genotypic and phenotypic data.

Results: TASUKE+ is equipped with the following functions. A Manhattan plot of a GWAS is displayed in parallel with polymorphism data corresponding to hundreds of whole genome sequences. A molecular phylogenetic tree can be reconstructed so that the relationships among the examined genomes are elucidated. In addition, users can create PCR primers near a prospective polymorphism.

PO0149: Methods: Bioinformatics

From Tropical Crop Breeding to Industrial Processing: Integrated Data Management with a Research Resource Planning

Tristan Duminil, Doriane SAS, Nice, France

Tropical crops intended for industrial purposes such as Cocoa, Coffee, Cannabis, Tobacco and Sugarcane require specific research and development to ensure the availability of quality production at the best cost. Complex research processes are involved, and huge amounts of data generated at every step of the transformation process, from plant breeding to sensorial analysis, including agronomical insights, raw material production and industrial processing. It's a key issue to integrate the whole research activity and take the good decisions to address the goals of the industrial organization.

Even more with the outcome of easily accessible sensor and environmental data bringing new potential information on condition to be able to collect, store and analyse it. Combined with the needs of the market, generated by an integrated chain of return of experience from the food industry and from end-consumers, it makes data management a real challenge for agro-industrial research centers.

Before using Research Resource Planning¹ (RRP), Food Industry organizations were lacking tools to analyse and get concrete indicators out of genotypic, phenotypic and processing data altogether, with restricted time and resources. Most of the historical data handling and analysis solutions combined various software or solutions implying complexity and difficulties for global, multidimensional analysis.

Are below presented the possibilities offered by an RRP such as RnDExperience® by Doriane® (www.doriane.com) to integrate data management tool for tropical crop research centers, and the benefits provided in terms of integrated analysis and decision making focusing on processing information and consumers expectations into industrial production strategies.

PE0150: Methods: Bioinformatics

Comparing Genomes and Visualizing Synteny with JBrowse 2

Robert Buels, Bioengineering, Berkeley, CA, **Colin Diesh**, University of California Berkeley, Bioengineering, Berkeley, CA, Garrett Stevens, University of California at Berkeley, Berkeley, CA, Ian Holmes, University of California, Berkeley, Berkeley, CA and JBrowse

Genome browsers provide a general framework for visualizing genome-mappable data, but they are traditionally limited to data from a single organism at a time. We created JBrowse 2 to show comparative visualizations of multiple genomes and structural variants relative to a reference genome, as well as enabling new visualization modalities that are not possible with JBrowse 1. JBrowse 2 can visualize multiple genomes on a single screen using linear views (c.f. MAUVE, GBrowseSyn) and dynamically-rotatable circular views (c.f. Circos), and can show connections between homologous features across different views (for highlighting synteny), or connections between reads that span breakpoints (arising from chromosomal rearrangements or other gross structural variants). JBrowse 2 is available as a self-contained desktop application (e.g. a single executable file) that does not require any web server setup, but (like JBrowse 1) it can also be deployed on a web server to publish genomic data to a wider community of researchers. We will describe a workflow to load synteny data from popular tools such as MCScanX, DAGChainer, or whole genome alignment pipelines and show the results in JBrowse 2.

PO0151: Methods: Bioinformatics

Affordable Genome Browsers for Non-Model Organisms on Cloud

Subhashini Srinivasan, Institute of Bioinformatics and Applied Biotechnology, Bangalore, Karnataka, India and **Saurabh Babanrao Whadgar**, Institute of Bioinformatics and Applied Biotechnology, Bangalore, India

The genome of every organism is unique and deserves all the bells and whistles to establish genotype-phenotype relationships. As the cost of data generation keeps plummeting the applications of genome sequencing have exploded in recent years to include studies of diverse plants and animals of interest to individuals investigators. However, the amount of data generated and information buried in them cannot be comprehended without looking for variations in the context of the respective reference genome and proteome. There is a need to collate reference genomes, proteomes, individual genomes and transcriptomes in a single framework to aid prioritization of trait-specific signatures by individual investigators for downstream validation and translation. We have integrated JBrowse with cloud architecture to make genome browsers both affordable and easily accessible to individual investigators for any organism of interest. Here, using an example, we will present our strategy to launch a browser within hours for costs as low as your cell phone bills per month.

PE0152: Methods: Bioinformatics

Breedbase: A Digital Ecosystem for Plant Breeding of Root and Tuber Crops and Beyond

Guillaume J. Bauchet, Boyce Thompson Institute, Ithaca, NY

More than 300 million people below the poverty line in developing countries depend on root, tuber and banana (RTB) crops for food and income. National and International research centers (ie: NARO Uganda, TARI Tanzania, CGIAR centers (CIP, IITA)) working on the RTB crops like cassava, sweet potato, yam, banana and plantain.

Crop breeding experiments are data intensive and long lasting, yielding thousands of phenotype and genotype datasets over years. Accurate breeding decision requires effective data management for collection, storage, analysis.

Breedbase instances (ie: cassavabase.org, sweetpotatobase.org) are web-based databases allowing storage and analysis of both phenotypic and genotypic data. Breedbase stores trait ontologies, crosses, pedigrees, images, genotyping and sequencing data. Decision support tools are provided for phenotyping, breeding management and trial analysis (ie: Genomic Selection).

To ensure data continuum, Breedbase supports barcode based digital data acquisition using PhenoApps (<http://phenoapps.org/>) such as Fieldbook for field data collection or Coordinate (genotype tissue sampling). Breedbase interfaces with the Breeding API (BrAPI) that defines standard data objects and methods for exchanging data which allows individual software components to talk to one another. Breedbase offers plant breeding communities a digital ecosystem to support their experimental activities.

PO0153: Methods: Bioinformatics

Graphical Data Visualization Tools in Phytozome

Joseph W. Carlson¹, Jeremy Phillips², Patrick Davidson¹ and David M. Goodstein², (1)DOE Joint Genome Institute, Berkeley, CA, (2)DOE Joint Genome Institute, Walnut Creek, CA

The Phytozome project (<https://phytozome.jgi.doe.gov/>) is the JGI's compendium of landplant assemblies, annotations and analyses consisting of genomes sequenced by JGI, by collaborators and community members. Our web infrastructure uses a combination of a customized databases architecture and community standard databases (CHADO, InterMine and BioMart) for providing the datastore to deliver data.

Data resulting from queries to our web services - especially from our conventional interfaces such as BioMart and InterMine - can be presented in tabular form, however a conventional listing of results often does not give the insight provided by a graphical display.

To this end, we have developed or incorporated graphical presentations of aspects of our data, including:

1. Information of metabolic pathways is presented as a graph of reactions in a pathway, showing the organization of reactions and the genes producing the enzymes in each reaction. The layout can be manipulated on the web site and the resulting view saved as an SVG graphic for reuse.
2. Clustergrammer (Fernandez,N.F. et al. Sci.Data 4:170151 (2017)) is a third-party tool for visualization of gene expression data. We have incorporated this framework for our expression data for the display of expression of individual list of genes or coexpressed genes. We have also coordinated it with the metabolic pathway viewer to show expression of gene associated with a specific pathway.
3. The phylogenetic relationships of genes in a specific family. This elucidates the degree of conservation of a particular protein within the evolutionary history.

PE0154: Methods: Bioinformatics

Assembly-Free, High-Density Linkage Map Construction from Whole Genome Sequencing Data

Kyle Fletcher and Richard Michelmore, The Genome Center, University of California, Davis, CA

We have implemented a method for construction of high-density linkage maps from raw sequencing reads without the need for a genome assembly. Conventionally, reads are aligned to an assembly to identify polymorphic sites that are then used to call the genotypes of segregating progeny. Typically, read-mapping and variant calling are

computational bottlenecks. Therefore, assembly-free linkage mapping would streamline map construction. Polymorphic markers can be identified through interrogation of whole genome shotgun reads of both parents that can be classified as single or multi-nucleotide variants. A genotype table can be rapidly generated by the presence or absence of these markers in whole genome shotgun data of progeny that can be analyzed by established tools for linkage mapping. We have applied this approach to whole-genome sequencing data of F₁ progeny derived from the out-crossing oomycete *Bremia lactucae*. Analysis identified over 17,000 markers which segregated in a near 1:1 manner. Linkage analysis using MSTmap placed 99% of the markers into 15 linkage groups. These linkage groups were colinear with the chromosome-scale *B. lactucae* reference assembly. This pipeline fills a niche not currently filled by other software solutions.

PO0155: Methods: Bioinformatics

Fast and Accurate Genome Assembly and Polish Tools for Third Generation Sequencing (TGS) Reads

Jiang Hu, GrandOmics Biosciences Co.,Ltd., Wuhan, China

The third-generation long-read sequencing, such as the PacBio single-molecule real-time (SMRT) sequencing and Oxford Nanopore sequencing (ONT), can overcome the challenges that are inherent to short-read and can resolve complex and repetitive genomic regions in genomes. However, the high error rate of both sequencing technologies requires software tools that can efficiently correct those errors in the long reads. These tools are essential to produce a high quality sequence assembly by correcting the sequence errors at the initial assembly step as well as at the consensus generation step from assembled contigs, which usually are cost-inhibitive, time-consuming, and computer-resource intensive. Therefore, we developed two effective software, NextDenovo, a string graph-based *de novo* assembler for long reads, and NextPolish, a fast and efficient genome polishing tool, to generate a high quality genome assembly with great efficiency. Compared with other existing tools with similar functions, both NextDenovo and NextPolish can produce more accurate and contiguity assembly with significantly less computing resources. Binary releases can be found at <https://github.com/Nextomics/NextDenovo> and <https://github.com/Nextomics/NextPolish>.

PE0156: Methods: Bioinformatics

Repbase: A Comprehensive Database of Eukaryotic Repeat Sequences and Transposons

Kenji K. Kojima and Weidong Bao, Genetic Information Research Institute, Cupertino, CA

Repbase is a comprehensive collection of representative repeat sequences in eukaryotic genomes. Since its first development in 1992, Repbase has been serving as a well-curated reference database fundamental for almost all eukaryotic genome sequence analyses. One main usage of Repbase is to mask repetitive sequences from the genome of interest with software such as Censor or RepeatMasker. Because most of the repetitive sequences originated from various transposons as well as integrated viruses, Repbase is also a fundamental source for the studies of transposons and their impact on the genome evolution. Transposons contribute to the birth of many functional DNA components, such as promoter, enhancer, or insulator, and thus, demands of mapping (not masking) of repeats are rising rapidly too.

One distinguished feature of Repbase is the detailed characterization/annotation of the repeat sequences that includes the classification into a particular transposon superfamily, the average identity/similarity of sequence copies to the consensus, and the reconstructed protein sequences that were originally encoded by their ancestral transposons.

We have transitioned Repbase to a community-supported subscription model; starting April 12, 2019, an active subscription is required to access Repbase. We offer subscriptions at both the individual and academic institutional level. Subscribers can access to the latest versions of Repbase (including RepeatMasker version), the contents of our online journal Repbase Reports, pre-repeat-masked genomes, and repeat maps. We also offer unlimited times of use of online Censor (homology-based search tool against Repbase) and RTclass1 (a classification tool for non-LTR retrotransposons) for subscribers.

PO0157: Methods: Bioinformatics

TARP: Transposable Element Assembly Remap Pipeline

Ethan Holleman and Howard M. Laten, Loyola University Chicago, Chicago, IL

Transposable elements (TEs) are a major component of plant genomes, constituting up to 80% of some genomes. Originally regarded as little more than genetic parasites, an increasing number of studies have shown these elements contribute significantly to the phenotypes of many important crop species. Therefore, accurate and up to date information on the locations of TEs in assembled plant genomes is essential for understanding their functional impacts. Plant genome assemblies continue to be improved and updated, but public TE databases often remain based on original draft assemblies. There is a need for computational methods to reliably and accurately locate and remap TEs from outdated to current assemblies.

Here, we present Transposable Element Assembly Remap Pipeline or TARP. TARP utilizes iterative global alignments of TE consensus sequences to locate known and novel elements in updated plant genome assemblies without searching for either one of these explicitly. This allows investigators interested in the TE content of updated plant assemblies to utilize previous TE libraries and data to inform their search instead of starting from scratch.

Additionally, using flanking sequence comparisons TARP can distinguish between novel and known elements in a new assembly even if there have been significant changes in the locations of previously identified elements. This allows investigators to reassign TE positions to the most up to date assemblies more easily in order to take full advantage of the latest genome data.

PE0158: Methods: Bioinformatics

Mining Microbial Genomes from Datasets on the Sequence Read Archive

Bhavya Papudeshi¹, Haley Leffler¹, Sruthi Ganapaneni¹, Sheri Sanders², Carrie Ganote¹ and Tom Doak¹, (1)Indiana University, Bloomington, IN, (2)National Center for Genome Analysis Support, Pervasive Technology Institute, Bloomington, IN

The declining costs of genome sequencing and growing amounts of genetic data has allowed the field of genomics to become more integrated with computational analysis. The use of high performance clusters (HPC) is necessary to compute the large amounts of data in genomic projects, however, many biologists lack background experience in working with HPC systems, which limits their ability to best address their research questions. The National Center of Genome Analysis Support (NCGAS) is an NSF-funded center that focuses on filling this need, by providing training as workshops, bioinformatics support on projects, and access to compute resources. As a byproduct of helping research projects, we develop open source workflows and make them available to the community. Here we present a developed workflow that will assist researchers in mining the Sequence Read Archive (SRA), to identify environments/datasets potentially containing genomes of interest, and identify their closely related genomes. As a proof of concept, we used two genomes to test the developed workflow, selected to ensure the flexibility of the workflow to generate results in formats amiable to further downstream analysis, based on the research question. The developed pipeline is made available through GitHub (<https://github.com/NCGAS/CEWiT-REU-Identifying-datasets-in-SRA-using-Jetstream>), and available as a pre-installed workflow on the XSEDE Jetstream cloud computing infrastructure.

PO0159: Methods: Bioinformatics

EASEL: An Integrated and Accessible Framework for the Annotation of Eukaryotic Reference Genomes

Jeremy Bennett¹, Susan L McEvoy¹, Sumaira Zaman², Peter Richter³ and Jill L. Wegrzyn³, (1)University of Connecticut, Storrs, CT, (2)Department of Computer Science and Engineering, University of Connecticut, Storrs, CT, (3)Department of Ecology and Evolutionary Biology, University of Connecticut, Storrs, CT

High throughput technologies have increased both the number and quality of eukaryote genomes. This increase in reference genomes, and their associated contiguity, is not yet met with efficient and intelligent workflows for the accurate detection of protein coding genes. The prediction of a true gene, and the correct translation initiation start site (TIS), remains a tremendous challenge, especially for non-models with minimal transcriptomic resources. The majority of pipelines utilize RNA-seq reads or pre-assembled transcripts to train (apply supervised or semi-supervised methods) Hidden Markov Models to predict gene structures for species with minimal or substantial existing genomic resources. These programs struggle with predicting less common gene structures (long introns, micro-exons), finding the preferred TIS location, and distinguishing pseudogenes. We present EASEL (Efficient, Accurate, Scalable Eukaryotic modeLS), a genome annotation tool that leverages deep learning, RNA folding, and functional annotations to enhance gene prediction accuracy. EASEL features a deep LSTM network that has the capability to learn species-specific patterns to predict non-canonical gene structures and train in reasonable time. Existing high quality alignments from BUSCO and related protein sources train the network on intron/exon parameters. The implicated genomic regions are further refined via unsupervised training with RNA folding and traditional consensus patterns to improve TIS detection. Predicted proteins are subject to filtering via functional analysis information from the program EnTAP (gene family and protein domain signatures). The pipeline is benchmarked for user friendliness, efficiency, completeness, and accuracy in the detection of protein coding gene models.

PE0160: Methods: Bioinformatics

Indices for NGS Data and Gene Expression Data Registered in Public Databases

Tazro Ohta, Takeru Nakazato and **Hidemasa Bono**, Database Center for Life Science, Mishima, Japan

In the integrated database project in Japan, Database Center for Life Science (DBCLS) has developed computational tools for searching huge amount of data archived in the public repository (<https://dbcls.rois.ac.jp/services-en.html>)

Meanwhile, DNA Data Bank of Japan (DDBJ) has archived and maintained data from the high-throughput sequencing platforms in the International Nucleotide Sequence Database Collaboration (INSDC) with NCBI Genbank and EBI ENA (<http://www.insdc.org/>). In collaboration with DDBJ, we made a search engine for metadata of these INSDC databases which consist of Bioproject, Biosample, and Sequence Read Archive (SRA), which is called DBCLS SRA (<http://sra.dbcls.jp/>). DBCLS SRA is now linked from DDBJ website, but it is planned to be used in DDBJ officially.

By the high-throughput sequencing platform, tens of thousands of RNA-seq data have been archived as transcriptome data in SRA. On the other hand, transcriptome data from microarray is still the majority of data in public gene expression databases known as NCBI Gene Expression Omnibus (GEO) and EBI ArrayExpress (AE). Furthermore, DNA DataBank of Japan started similar repository called Genomic Expression Archive (GEA) in 2018. Thus, it is not easy to draw new discoveries by comparing datasets from those transcriptomes because of the complexity of relationships among those databases. We constructed an index for those gene expression data repositories, called all of gene expression (AOE) to integrate publicly available gene expression data (GEO, AE and GEA). The web interface of AOE (<https://aoe.dbcls.jp/>) can graphically query data in addition to the application programming interface. By collecting gene expression data by RNA-seq from SRA, AOE also includes data not included in GEO, AE and GEA.

Both DBCLS SRA and AOE are freely available without any registration.

PO0161: Methods: Bioinformatics

PpsPCP: A Plant Presence/Absence Variants Scanner and Pan-Genome Construction Pipeline

Muhammad Tahir ul Qamar, Xitong Zhu, Feng Xing, Jiaming Song and Ling-Ling Chen, Huazhong Agricultural University, Wuhan, China

The recent advancement in plant genomics research has been rapidly developed which results in large volume of data submission in public databases and enable data acquisition, whole-genome sequence data analyses and comparative genomic analyses through different pipelines. The intra-species genetic variations, especially in the

form of presence/absence variants (PAVs) are a key to natural and artificial selection. Whole genome re-sequencing data is usually mapped to a reference genome when genomes of different individuals are compared. However, one single reference genome is insufficient to epitomize genetic makeup of different individuals in the same species and often ignore some important genes and leads to inaccurate estimation of genetic diversity. To get a comprehensive map of genetic variations, phenotypic variations and genomic diversity, it is crucial to construct a pan-genome including all the specific PAVs. Since the idea of pan-genomics emerged several tools and pipelines have been introduced for prokaryotic pan-genomics. However, not a single comprehensive pipeline has been reported which could overcome multiple challenges associated with eukaryotic pan-genomics. To aid the eukaryotic pan-genomic studies, I developed ppsPCP, a novel pipeline which takes advantage of assembled plant genomes, screen PAVs from them and develops a completely annotated pan-genome. This ppsPCP pipeline is benchmarked with model cereal species rice and model dicot species *Arabidopsis thaliana*. In case of rice, ppsPCP constructed a 420 MB sized pan genome containing 43,082 genes. A total of 11,677 PAVs and 4,213 genes were screened and added to the rice pan-genome. In case of *A. thaliana*, ppsPCP constructed a 122 MB sized pan-genome containing 34,899 genes. A total of 7,480 PAVs and 1,432 genes were screened and added to the *A. thaliana* pan-genome. Furthermore, to evaluate the quality of developed pan genomes by ppsPCP, we compared our rice results with recently reported pan-genome developed from 3,010 diverse accessions of Asian cultivated rice and *A. thaliana* results with pan-genome of 19 ecotypes. In case of rice, only 2,650 genes out of 50,955 genes were found to be different, and 48,305 genes were fully mapped to ppsPCP constructed pan genome. While, in case of *A. thaliana*, the constructed pan-genome of ppsPCP contains 1,687 more genes compared to the pan-genome. All the input and output data can be downloaded from ppsPCP webpage. We believe with these unique features of PAV scanning and building a pan-genome together with its annotation, ppsPCP will be useful for plant pan-genomic studies and aid researchers to study genetic/phenotypic variations and genomic diversity.

Availability and implementation:

The ppsPCP is freely available at github DOI: <https://github.com/Zhuxitong/ppsPCP> and webpage <http://cbi.hzau.edu.cn/ppsPCP/>.

PE0162: Methods: Bioinformatics

Determining Effective Dimensionality Reduction Tools for Genetic Data used in Deep Learning Models

Therin Young¹, Zihao Zheng², Koushik Nagasubramanian³, Patrick S. Schnable² and Baskar Ganapathysubramanian⁴, (1)Iowa State University, Ames, IA, (2)Department of Agronomy, Iowa State University, Ames, IA, (3)Department of Electrical Engineering, Iowa State University, Ames, IA, (4)Department of Mechanical Engineering, Iowa State University, Ames, IA

The high dimensionality of genome data poses significant challenges (overfitting, high correlation among input data, need for multiple genotypes for training well, etc.) for applying machine learning models. Hence, there is increasing interest in reducing the dimensionality of the input data (i.e. SNP data) before feeding into deep learning models. In this study, we compared a variety of data representation methods for deep learning-based dimensionality reduction of genomic data. The reconstruction error of the genomic data was assessed to determine the best data representation for deep learning-based dimensionality reduction. In addition to autoencoders, the effectiveness of the data representations methods was assessed for classical dimensionality reduction methods such as PCA. This study can serve as a practitioner's guide to effective genomic data representation for deep learning applications.

PO0163: Methods: Bioinformatics

Fast Parallelized Sampling of Bayesian Linear Mixed Models for Whole-Genome Prediction

Tianjing Zhao, University of California, Davis, Davis, CA, Rohan L. Fernando, Department of Animal Science, Iowa State University, Ames, IA, Dorian Garrick, Massey University, Ruakura, New Zealand and Hao Cheng, Department of Animal Science, University of California, Davis, Davis, CA

Bayesian linear mixed models (BLMM) are widely used in genomic prediction, where the effects of all markers are estimated simultaneously combining the information from the phenotypic data and priors for the marker effects.

Inferences from most Bayesian regression methods are based on Markov chain Monte Carlo methods, where statistics are computed from a Markov chain constructed to have a stationary distribution equal to the posterior distribution of the unknown parameters. In practice, chains of tens of thousands steps are typically used in whole-genome Bayesian analyses, which is computationally intensive. We have proposed a fast parallelized algorithm for BLMM called independent intensive BLMM (II-BLMM, "II" stands for "parallel") and shown how the sampling of marker effects can be made independent within each step of the chain. This is done by augmenting the marker covariate matrix with the number p of markers new rows such that columns of the augmented marker covariate matrix are orthogonal. Ideally, the computations at each step of the MCMC chain can be accelerated by the number k of computer processors up to the number p of markers. We demonstrate II-BLMM algorithm using the prior for BayesCpi, a Bayesian variable selection regression method, applied to simulated data with 50,000 individuals and a medium-density marker panel (~ 50,000 markers). To reach about the same accuracy as the conventional samplers for BayesCpi required less than 30 minutes using II-BLMM algorithm on 24 nodes with 24 cores on each node. In this case, II-BLMM algorithm required one tenth of the computation time of conventional samplers for BayesCpi. Addressing the heavy computational burden associated with Bayesian methods by parallel computing will lead to greater use of these methods.

PE0164: Methods: Bioinformatics

Approximate Bayesian Computational Statistical Methods to Identify Loci Under Selection from Yeast Genomic Data

Martyna Lukaszewicz, University of Idaho, Moscow, ID and University of Idaho Yeast Project Authors

Genomic data provides the possibility to learn how the environmental conditions structure the genetic make-up of organisms. However, analytical tools have not kept up with the wealth of genomic data, so that we are still limited in our ability to make detailed inferences about how natural selection and other evolutionary processes affect genomic variation. We are using genomic data from experimental evolution in yeast (*Saccharomyces cerevisiae*) to observe the effect of interactions between selection, migration, populations and recombination rate of parental genome, and validate novel statistical tools. We are testing the hypothesis that weaker selection and higher migration and recombination rates reduce the size of genomic islands. We created an admixed ancestral population and imposed divergent selection, under controlled conditions of recombination and migration rates. We developed a simulator to simulate genomic data under divergent selection with migration. The simulator is used to develop approximate Bayesian computational (ABC) methods to make inferences on population genetic parameters. We are determining which population genetics statistics are the most informative for ABC estimation of parameters and identification of selected loci. We will validate these ABC methods on the empirical yeast dataset.

PO0165: Methods: Bioinformatics

Supervised Classification of Plant Single Cell RNA-Seq Data

Song Li, Virginia Tech, Blacksburg, VA

Single cell RNA-seq technologies have recently been applied in profiling the transcriptome of Arabidopsis root and have generated several large-scale expression data sets. An important step for scRNA-seq analysis is to assign cell types to the different cell populations. Current approaches include generating an index of cell identity scores, performing correlation analysis with bulk tissue RNA-seq data or inspecting expression patterns of known marker genes. In this work, we test several machine learning methods to automatically identify cell types in scRNA-seq data of Arabidopsis root. Our pipeline includes conventional classification approaches such as k-nearest neighbors, random forest (RF) and support vector machine (SVM). We also tested multi-task, deep neural networks (NN) with conventional, contrastive or triplet loss-functions. In our comparisons, random forest has achieved the best average performance (mean Average Precision, mAP > 0.85). Interestingly, any specific machine learning method can perform better in some cell types but not in others. For example, NN-based classifiers performed best in xylem and trichoblast cells, whereas SVM and RF performed best in endodermis and phloem cells. Regardless of which machine learning methods were used, endodermis and phloem cell achieved best performance, suggesting these two cell types have expression patterns that are distinctively different from other cell types.

PE0166: Methods: Bioinformatics

Visualization of Trait Ontology Having a Large Number of Traits

Daisuke Horyu, Institute of Crop Science, NARO, Tsukuba, Ibaraki, Japan, Akio Onogi, National Agriculture and Food Research Organization, Tsukuba, Japan and Takeshi Hayashi, Institute of Crop Science, NARO, Tsukuba, Japan

Even though various advanced measuring devices become popular in these days, phenotyping of crops still consumes much time and labor in many cases. Therefore, for the crop breeding organizations, it is an important issue to reuse already collected phenotyping data. Our organization is also working on construction of breeding database. Our soybean breeding database is one of such databases, and it has around 550 traits. When the user intends to use the database holding such a large number of traits, the problem that it is difficult to know if the traits similar to the traits that the user interested in exists in the database arise. Constructing a trait ontology is one of the methods to solve this difficulty. However, not only constructing a trait ontology, but we have to think about how to implement the ontology into the retrieval system of the breeding database. Making a system that enables to look out over the traits included in the database through visualization of trait ontology, and to implement it as an element of the database retrieval interface is one of the solutions for this issue. Considering the implementation on the web-based retrieval interface, we must construct it on the JavaScript. Taking these circumstances into account, we have tested the capability of JavaScript based network visualization libraries, such as Cytoscape.js, for visualizing large scaled trait ontology. The obtained images can greatly help understanding the relationships between traits.

PO0167: Methods: Bioinformatics

simplePHENOTYPES: Simulation of Pleiotropic, Linked and Epistatic Phenotypes in R

Samuel B Fernandes, University of Illinois at Urbana-Champaign, Urbana, IL and Alexander E. Lipka, University of Illinois, Urbana, IL

The potential for unprecedented study of genomic sources underlying phenotypic variability is facilitated by recent advances in genotyping and phenotyping techniques. It is now possible to deploy powerful statistical methods such as multivariate analysis to the data produced by these techniques, which could facilitate the identification of genomic regions likely to harbor pleiotropic loci. As the demand for multivariate analyses increases, it is imperative that optimal tools are available to validate and compare different implementations of these approaches. Therefore, we created an R package called simplePHENOTYPES that simulates traits controlled by pleiotropic, pseudo-pleiotropic, and non-pleiotropic quantitative trait nucleotides (QTNs). With user-inputted marker data, simplePHENOTYPES will simulate phenotypes for a specific number of correlated traits controlled by QTNs with additive, dominance, or additive x additive epistatic contributions. To account for differential linkage disequilibrium and minor allele frequencies across experiments, we include the option for selecting a different set of QTNs in each replication. Non-genetic phenotypic variability is simulated assuming uncorrelated multivariate normally distributed residuals, where the residual variances are proportional to the desired heritabilities. The user has the option of assigning specific heritabilities for the, as well as assigning the pairwise correlation between the traits. Finally, output files are saved in an analysis-ready format for different software including PLINK, GEMMA, TASSEL and GAPIT.

PE0168: Methods: Bioinformatics

Field Phenotyping with Natural Language

Colleen Yanarella, Ian Braun and Carolyn J. Lawrence-Dill, Iowa State University, Ames, IA

Computational linguistics techniques stand to revolutionize how researchers document and compute on phenotypic descriptions. Here we propose to use speech recognition along with Natural Language Processing (NLP) and Machine Learning (ML) for speech-to-text reporting of spoken phenotypic descriptions. We hypothesize that datasets generated in this way can be used for association genetics (and various other applications) irrespective of whether and how those descriptions adhere to current standards for trait data collection. To conduct this investigation, we will use unique two-dimensional (2D) barcodes that will be created for each maize plant in the field. A mobile cellular device application will be developed to scan each barcode and will be used to collect spoken phenotypic descriptions (including measurements) for each plant. Computational methods to determine semantic similarity among phenotypic descriptions will be applied (see Braun and Lawrence-Dill at <https://doi.org/10.1101/689976> for methods detail). Using these methods, we will crowdsource phenotypic

descriptions and compare the performance of a group of experts to a group of non-expert citizen scientists to determine whether subject matter expertise is necessary for biologically meaningful semantic similarity networks.

PO0169: Methods: Bioinformatics

Direct Computation on Phenotypic Descriptions for Novel Candidate Gene Prediction

Ian Braun and Carolyn J. Lawrence-Dill, Iowa State University, Ames, IA

Natural language descriptions of plant phenotypes present in databases and the scientific literature are a rich source of information for biological research that seeks to untangle relationships between genes and observable phenotypes, such as plant health or size. The volume and unstructured nature of these text descriptions however necessitates a computational approach for leveraging them to predict gene-to-phenotype associations. We computationally translated descriptions of plant phenotypes into structured representations that can be processed to identify biologically meaningful associations. These representations include the EQ (Entity-Quality) formalism, which uses terms from biological ontologies to represent phenotypes in a standardized, semantically-rich format. Ontology terms are mapped to text descriptions with a combination of string-matching and word embedding algorithms. Our computationally produced representations of text descriptions also include numerical vectors, generated using either a bag-of-words approach or document embedding. We compared resulting phenotype similarity measures to those derived from manually curated data to determine the performance of each method. Computationally derived EQ and vector representations were comparably successful in recapitulating biological truth to representations created through manual EQ statement curation. Moreover, these computational methods for generating representations of phenotypes are scalable to large quantities of text because they require no human input. These results indicate that it is now possible to computationally and automatically produce and populate large-scale information resources that enable researchers to query phenotypic descriptions directly. Ongoing work to produce phenomics-focused text mining tools for the bioinformatics community and a resource for exploring the results of this work for the plant biology community is discussed.

PE0170: Methods: Bioinformatics

Bivariate Genomic Selection for Boosting Predictabilities of Agronomically Important Traits

Rui Zhou, university of california, riverside, riverside, CA

The multivariate genomic selection (GS) models have not been adequately studied and their potential advantages over univariate GS models remain unclear. In this study, we developed a bivariate (2D) BLUP GS model and demonstrated its significant advantages over the univariate (1D) BLUP GS model with a rice dataset by analyzing 4 traditional traits, i.e., YIELD, KGW (1000-grain weight), GRAIN (grain number), and TILLER (tiller), as well as 1000 metabolomic traits. HAT methodology has been incorporated in the 2D BLUP GS model to increase computational efficiency by avoiding conventional cross-validation. The results indicated that (1) the 2D BLUP GS analysis generally produces higher predictabilities for two traits than those achieved by the analysis of individual traits using 1D BLUP GS model, and (2) selected metabolites may be utilized as ancillary traits in the 2D BLUP GS analysis to further boost the predictability of traditional traits, especially for agronomically important traits with low 1D predictabilities.

PO0171: Methods: Bioinformatics

Prediction of Transcription Factor Binding Sites Based on Network Data

Janik Sielemann and Andrea Braeutigam, Bielefeld University, Bielefeld, Germany

Regulation of transcription is a key factor for dynamic adaptation to variable environmental conditions. The binding of transcription factors to *cis*-elements plays a leading role in transcriptional regulation in plants. However, most transcription factors cannot be assigned to their specific binding motif yet. RNA-seq data of *Arabidopsis thaliana* was used to predict gene regulatory networks with a machine learning algorithm. DAP-seq data showed that binding motifs of transcription factors enriched in promoters of target genes scoring high in the predicted networks in 26% of the testable cases. Thus, motif prediction is likely possible for a subset of the transcription factors. The predicted target genes of 2,976 transcription factors were used as the basis for motif prediction.

In addition to the 3 kb upstream region of target genes, open chromatin regions were used to predict binding motifs. These regions were determined based on publicly available assay for transposase accessible chromatin data. Both, the plain 3 kb upstream region of predicted target genes and open chromatin regions close to target genes, revealed sets of putatively targeted sequences for all transcription factors in *Arabidopsis thaliana*. The motif prediction was calculated using motif-based sequence analysis tools (MEME-suite) and predicted motifs were compared to published motifs obtained from the JASPAR database. Prediction of binding motifs based on the accessible 3 kb upstream region performed better than random. The analysis indicates that the consideration of open regions improves the prediction accuracy of target motifs in *A. thaliana*. Since accuracy was limited, transcription factor family information was stacked to obtain higher confidence predictions.

PE0172: Methods: Bioinformatics

MOMS: A Chromosome-Level Genome Scaffolder Using Multi-Channel Optical Maps

Jiang Xu¹, Baosheng Liao², Shuiming Xiao², Shuai Guo², Xiaofeng Shen², Yang Chu¹, Wenguang Wu², **Qiushi Li¹** and Shilin Chen², (1)China Academy of Chinese Medical Sciences, BeiJing, China, (2)Institute of Chinese Materia Medica, China Academy of Chinese Medical Sciences, Beijing, China

Optical mapping is a fundamental tool for genome assembly and has been widely used in modern genome projects. Here we report a multi-channel optical map scaffolder (MOMS), with uncapped enzymes tolerance in chromosome-level optical map scaffolding, which imparts limitless potency to the optical maps-based scaffolding approach.

To facilitate the coordinate transformation under different channels and avoid error propagation, an original data structure-directed node graph (DNG) is designed for representing the association between optical maps under different channels and linkage between adjacent genomic regions. We also employed a heuristic algorithm to mine the path traversal and resolve conflicts. In the gap-filling stage, we combined the gap-aware mapping and multiple restriction enzyme site alignment. We comprised MOMaligner and a SSPACE translator to scaffold the genomic sequences.

For NA12878 human genome, the results showed that MOMS significantly improved the contiguity and completeness of the initial assembly to scaffold N50 ~90 Mbp, incorporating more data compared to the standard hybrid scaffolding with Bionano Solve. Most of the filled gaps (48/57, 84.2%) were validated.

We also used MOMS to finish three additional genomes with high contiguity (N50 >10 Mbp) at low expense. We found that a chromosomally dispersed gene family (olfactory receptors) expanded in ruminants and cluster based tandem replications may contribute to this expansion.

PO0173: Methods: Bioinformatics

Platanus-allee is a *de novo* Haplotype Assembler Designed for Organisms with Highly Divergent Haplotypes: An Update

Rei Kajitani¹, Miki Okuno¹, Hiroyuki Tanaka¹, Atsushi Toyoda² and Takehiko Itoh¹, (1)School of Life Science and Technology/Tokyo Institute of technology/Japan, Tokyo, Japan, (2)Advanced Genomics Center, NIG, ROIS, Shizuoka, Japan

Highly divergent genomic regions (HDRs) within species are associated with fundamental biological phenomena such as speciation, variable morphs, and sex-determination. These regions are possibly inaccessible with the standard read mapping-first methods due to extreme differences between the haplotypes (>5%). We developed a versatile haplotype assembler, Platanus-allee (<http://platanus.bio.titech.ac.jp/platanus2>), which constructed each haplotype independently using iterative simplifications of the assembly graphs and the haplotype synteny-based correction. Although it requires one shot-reads library, it can also utilize various libraries such as mate-pairs, single-molecule long-reads and 10x linked-reads.

Benchmarks were performed for data acquisition from diploid organisms whose heterozygosities ranged from 0.1 to >3 %. The tools benchmarked included a long-read-based one (FALCON-Unzip) and a 10x linked-reads-based one

(Supernova). Recalls and precisions of the resultant haplotypes were evaluated using the reference genome or the synthetic long-reads. Platanus-allee accurately assembled a large range of haplotypes, especially for HDRs. Remarkably, HDRs were detected for all the organisms including humans. The human HDRs consisted of major histocompatibility complex (MHC) regions and ones absent from the reference genome. In addition, Platanus-allee was effective in constructing conventional pseudo-haploid (consensus) genomes for the numerous highly heterozygous organisms.

Platanus-allee was first published in 2019 and with the recent update, the misassemblies were reduced to less than a half in most of the cases. Further, the updates reduced gap-rates and the run-times for all the cases. In summary, we describe the Platanus-allee latest updates, information, and application examples for practical genome projects including a combination of the other tools and long-reads.

PE0174: Methods: Bioinformatics

Identifying Parental Alleles with Bionano Genomics' Ultra-High Molecular Weight DNA

Joyce Lee, Andy Wing Chun Pang, Thomas Anantharaman, Alex R. Hastie and Mark Oldakowski, Bionano Genomics, San Diego, CA

Complete diploid assemblies with correct phasing has always been a challenge with non-trivial applications in understanding diseases and reference building. Trio-binning of proband reads using parental reads has been applied recently to partition proband reads prior to the assembly of the binned paternal and maternal sets (Koren et al., 2019). With the advantage of the Bionano Genomics ultra-high molecular weight DNA spanning complex genomic regions, such as segmental duplications and repeat expansions, trio-binning of proband Bionano reads to the Bionano parental haplotype-aware assemblies gives contiguous assembly of each allele.

Using a human trio dataset, proband molecules that uniquely aligned to either of the parents were assigned to their respective maternal or paternal bins. Other molecules were evenly distributed to the bins. Following binning, each of the bins assembled to about 2.95 Gbp with N50s of about 65 Mbp, and with less than 110 loci showing two alleles, where potentially both parents share the same alleles. To further segregate the alleles, a cross-checking step was performed by aligning the proband's binned assemblies to the parents' assemblies to identify regions where both parents shared the same allele, but homozygous in one and heterozygous in the other. We can then eliminate one of the parents' heterozygous alleles and improve the parental anchors for the next round of trio-binning. We have applied this procedure to trio-binning of several animal genomes as well.

In conclusion, trio-binning with Bionano Genomics' ultra-long DNA molecules can efficiently segregate haplotypes and assemble each of the alleles with superior contiguity.

PO0175: Methods: Bioinformatics

kWIP: The k-mer Weighted Inner Product, a *de novo* Estimator of Genetic Similarity

Kevin D. Murray¹, Christfried Webers², Cheng Soon Ong², Justin O. Borevitz¹ and **Norman Warthmann¹**, (1)The Australian National University, Canberra, ACT, Australia, (2)Data61/CSIRO, Canberra, ACT, Australia

Rapidly estimating genetic relatedness of individuals (or "samples") directly from sequencing data is often desired, e.g., to establish sample identity and detect mix-up, to uncover non-obvious genomic variation, or to assess population structure without a reference genome and the associated biases. Determining genetic relatedness directly from high-throughput DNA sequencing data, quickly and in an unbiased manner, demands novel, computationally efficient methods.

We present the *k*-mer Weighted Inner Product (kWIP), an assembly-, and alignment-free estimator of genetic similarity along with ready-to-use software. kWIP combines a probabilistic data structure with a novel metric, the weighted inner product (WIP), to efficiently calculate pairwise similarity between sequencing runs from their *k*-mer counts weighted by their information content. The kWIP software produces a distance matrix, which can be further

analysed and visualised. Our method does not require prior knowledge of the underlying genomes. kWIP is written in C++, licensed under the GNU GPL, and is available from <https://github.com/kdmurray91/kwip>.

PE0176: Methods: Bioinformatics

Computing Pipeline for Genomic Prediction and Estimation Using Haplotypes and SNP Markers

Dzianis Prakapenka and Yang Da, Department of Animal Science, University of Minnesota, Saint Paul, MN

The haplotype analysis for genomic prediction and estimation requires considerably more data processing and has many more possible configurations of the prediction model than single-SNP analysis. To facilitate haplotype analysis for genomic prediction and estimation, we developed a computing pipeline to implement haplotype analysis. The pipeline includes three components, preparation of input data for haplotype analysis, genomic prediction and estimation using GVCHAP, and analysis of GVCHAP results. The input preparation starts with formatting SNP data for two imputing programs. Based on the imputed haplotypes, a utility program defines haplotype blocks by a fixed number of SNPs or a fixed distance in base pairs per block, where each block is treated as a multi-allelic locus and is formatted as haplotype genotypes with two haplotypes per genotype. The haplotype genotypes are used as an input file for running GVCHAP. Another utility program fills in most of the parameter file required by GVCHAP as an input file. The data preparation step also contains utility programs for defining validation samples by random assignment of individuals to each validation sample or by a user provided list of individuals for assigning to validation samples. GVCHAP is the main program for genomic prediction and estimation providing GREML estimates and GBLUP for additive and dominance effects of haplotypes and single SNPs. To reduce the computing time in cross validations due to calculation of genomic relationships, GVCBLUP has a 2-step strategy to save the genomic relationship matrix during the first fold of validation and read in the genomic relationships for the remaining folds of validations. This 2-step strategy is helpful for k-fold validations and for multiple traits. The last component of the computing pipeline calculates observed prediction accuracies and produce input file for graphical analysis of haplotype and SNP heritabilities.

PO0177: Methods: Bioinformatics

SyRI: Finding Genomic Rearrangements and Local Sequence Differences from Whole-Genome Assemblies

Manish Goel, Max Planck Institute for Plant Breeding Research, Cologne, Germany

Genomic differences range from single nucleotide differences to complex structural variation. Current methods typically annotate only differences in sequence (ranging from SNPs to large indels) but do not consider the full complexity of structural rearrangements (including inversions, translocations and duplications), where highly similar sequences change in location, orientation, or copy-number. These rearrangements have been reported to be associated with several biological differences between organisms, however they are still scantily studied.

Here we present SyRI, a pairwise whole-genome comparison tool for chromosome-level assemblies. SyRI uses a unique approach where it first identifies all syntenic (non-rearranged) regions between two genomes. Since all non-syntenic regions are structural rearrangements by definition, this transforms the difficult problem of rearrangement identification to a comparatively easier problem of rearrangement classification. After finding rearranged regions, SyRI searches for differences in the sequences, which are distinguished for residing in syntenic or rearranged regions. This distinction is important, as rearranged regions (and sequence differences within them) are inherited differently compared to syntenic regions. Using SyRI, we successfully identified rearrangements in human, *A. thaliana*, yeast, fruit fly, and maize genomes. Further, we also validate 92% (108/117) predicted translocations in *A. thaliana* using genetics.

PE0178: Methods: Bioinformatics

SMAP: A Versatile Approach to Read-Backed Haplotyping in Stacked NGS Read Data

Tom Ruttink¹, Elisabeth Veeckman¹, Dries Schaumont¹, Sabine Van Glabeke¹, Annelies Haegeman¹, Hilde Muylle¹, Isabel Roldán-Ruiz¹, Jean-Paul Sampoux², Philippe Barre², Stephen Byrne³, Jie Kang⁴, Felix Van der Jeugt⁵ and Peter Dawyndt⁵, (1)ILVO Plant Sciences, Melle, Belgium, (2)INRA UR4 (URP3F), Lusignan, France, (3)Teagasc, Crop Science Department, Carlow, Ireland, (4)AG Research, Dunedin, New Zealand, (5)UGent, Dept. Applied mathematics, computer science and statistics, Gent, Belgium

We have developed a computational algorithm called "SMAP" (short for Stack Mapping Anchor Points) for read-backed haplotyping in stacked NGS reads, applicable to highly multiplex amplicon sequencing (HiPlex) or Genotyping-by-Sequencing (GBS) data. The method first delineates read stacks on the reference genome taking into account read mapping position polymorphisms. This module is very useful to validate all read preprocessing steps, including read trimming and read mapping, and provides graphical summaries of read distribution per sample and across the sample set, such as number of stacks, completeness, and read depth saturation curves. Combined with SNP polymorphisms in those stacks, the novel method subsequently creates haplotypes using read-backed phasing and estimates the relative haplotype frequency per stack, across all samples. Various user-controlled filter criteria can be defined on the command line. Using multi-allelic haplotypes instead of bi-allelic SNPs reduces redundancy in marker sets, while increasing the discriminative power per molecular marker. The method yields quantitative haplotype frequencies for pool-Seq data or discrete haplotype calls for diploid or tetraploid individuals. This approach holds great promise for the use of allele frequency fingerprinting of heterogeneous plant populations of outbreeding species in the context of variety identification, parental analysis, association mapping, genomic selection, or characterization of genetic resources. We are currently working on expanding the method to high-throughput characterization of CRISPR-mediated genome edited lines.

PO0179: Methods: Bioinformatics

FAANGMine: Genomic Data Mining Tools for Domesticated Animal Species

Deborah A. Triant¹, Justin J. Le Tourneau¹, Md Shamimuzzaman¹, Amy T. Walsh¹, Darren Hagen² and Christine G Elsik¹, (1)University of Missouri, Columbia, MO, (2)Oklahoma State University, Stillwater, OK

FAANGMine (<http://faangmine.org>) is a genomic data mining warehouse for domesticated animal species, including species of interest to the Functional Annotation of Animal Genomes (FAANG) Consortium. FAANGMine provides simple and sophisticated search tools to enable researchers without scripting skills to create and export customized annotation datasets merged with their own research data for use in downstream analyses. The first FAANGMine release (v1.1) contains genomes of cat, chicken, cow, dog, goat, horse, pig, sheep and water buffalo. Gene annotations of human, mouse and rat are also included to facilitate comparison to model organisms. FAANGMine uses the InterMine platform to integrate data from external sources, including reference genome assemblies, genes (NCBI, Ensembl), proteins (UniProt), protein families and domains (InterPro), orthologs and paralogs (EnsemblCompara, OrthoDB, TreeFam), pathways (KEGG, Reactome), interactions (BioGRID, IntAct), Gene Ontology (GO), QTL (AnimalQTLdb), variation (Ensembl) and publications (PubMed). Built-in query templates provide starting points for data exploration, while the QueryBuilder tool supports construction of complex queries. The List Analysis and Genomic Regions search tools execute queries based on uploaded lists of identifiers and genome coordinates, respectively. Data can be exported in a variety of formats, including gff, fasta, json and tab-delimited files. We are working to expand FAANGMine by incorporating data generated by the FAANG Consortium to enable fine-grained data mining of functional elements in combination with gene annotations and additional biological data.

PE0180: Methods: Bioinformatics

An Update on Database Growth and Improvements of Animal QTLdb and CorrDB

Zhiliang Hu, Carissa Park and James M. Reecy, Department of Animal Science, Iowa State University, Ames, IA

The Animal Quantitative Trait Loci (QTL) Database and Animal Trait Correlation Database (CorrDB) have undergone phenomenal data growth and the addition of multiple database tools over the past few years. These

efforts have been well-received by the community, as evidenced by the increasing number of database web hits and data downloads every year. To date, there have been 174,241 QTL/associations on 2,062 traits from 6 species reported in 2,214 publications, 18,049 correlations on 629 traits, and 3,083 heritability data on 912 traits reported in 374 publications curated into the databases. The Animal QTLdb and/or CorrDB have been cited by over 1,000 scientific journal papers. New developments since our last update include (1) a more streamlined curation workflow and tools to support collaborative data curation between different curators; (2) more links between data types available for public access, including epistatic and pleiotropic QTL/associations, related data from follow-up studies, combined analyses of multiple studies, re-analyses, confirmation studies, or additional reports from the same experiments; (3) an improved QTL/association enrichment tool that allows user-selected genome regions and trait types to be compared against the rest of the genome to identify possible areas of over-representation; and (4) use of InterMine to facilitate genome-based QTL/association and correlation data linking, searches, and downloads. Future developments of the databases will include support of multiple genome builds within a species, and tools for liftover of QTL/association mapping data across different versions of genome assemblies.

PO0181: Methods: Bioinformatics

Development of a R Package (GALLO) for QTL and Gene Annotation in Livestock Species: A Case Study using Positional Candidate Genes for Fertility Traits in Beef Cattle

Pablo A.S. Fonseca¹, Aroa Suarez-Vega¹, Gabriele Marras² and Angela Cánovas¹, (1)Centre for Genetic Improvement of Livestock, Department of Animal Biosciences, University of Guelph, Guelph, ON, Canada, (2)The Semex Alliance, Guelph, ON, Canada

Genomic Annotation in Livestock for positional candidate LOci (GALLO) is an R package, for the annotation of genes and QTLs located in genomic candidate regions, allowing the graphical visualization, data comparison among factors (i.e. methods, breeds, tissues, statistical models, studies, etc.), and QTL enrichment in different livestock species including cattle, pigs, sheep and chicken, among others. Here, we present a practical example of GALLO usage throughout analyzing candidate regions associated with scrotal circumference related traits (SCT) and spermatid related traits (ST) in six beef populations (two Brahman, two Nellore, one Canchim and one Tropical Composite). In general, a higher number of positional candidate genes were shared between SCT, even when compared with ST, independently of the study and/or breed. For example, age at scrotal circumference of 26 cm and percentage of normal sperm (PNS) shared 100% of the positional candidate genes, while the highest proportion of shared genes between ST evaluated in different studies was 57% between PNS and semen volume. The higher heritability and homogeneity of SC phenotypes when compared with ST corroborate the results. Additionally, GALLO shows that the genomic regions identified are enriched for fertility related QTLs. The integration of multiple studies allowed the identification of biological patterns that would not be possible to be identified analyzing the datasets individually. GALLO is a useful package for gene and QTL annotation, identification of hidden patterns across datasets, datamining of previous reported associations, and the efficient scrutinization of the genetic architecture of complex traits in livestock.

PE0182: Methods: Bioinformatics

Genetic Variation in the Holstein Bovine MHC

Rodrigo P. Savegnago¹, Gonzalo Rincon² and Cedric Gondro¹, (1)Michigan State University, East Lansing, MI, (2)Zoetis, Kalamazoo, MI

The major histocompatibility complex (MHC) is a group of highly polymorphic genes involved in the immunological defense of organisms. In cattle it is located on chromosome 23 and known as bovine lymphocyte antigen (BoLA). Knowledge of BoLA variants has practical applications for the design of vaccines and to develop mating strategies to generate offspring with desirable haplotypes. In this study we quantified the number of SNP segregating in the BoLA region and the BoLA alleles using RNA-seq obtained from blood samples of 182 Holstein cows. The analytical workflow consisted of trimming the read adapters with Trimmomatic, read mapping onto the bovine reference genome assembly ARS-UCD1.2 with STAR, marking duplicated reads using Picard, SplitNCigarReads and GATK's Base Quality Score Recalibration, followed by variant calling with bcftools. VCFtools was then used to filter each variant in the VCF file for a SNP quality score ≥ 20 , minimum depth of 10, MAF of 0.01 and a missing data cutoff threshold of 0.5. Across the entire genome there were 91,797 SNP and 7%

(6,217) of these were on chromosome 23. In the BoLA region, which encompasses 4.51% of the bovine genome, we found 4,142 SNP which amounted to 67% of those found on chromosome 23. The IPD-MHC database has 572 BoLA alleles cataloged; 73 of these were detected in our samples (37 alleles of 8 BoLA class I genes and 36 alleles of 3 BoLA class II genes). The average MAF in the BoLA region was 0.18.

PO0183: Methods: Bioinformatics

Exact Distribution of Linkage Disequilibrium in the Presence of Mutation, Selection or Minor Allele Frequency Filtering

Jiayi Qu, University of California, Davis, DAVIS, CA, Stephen D. Kachman, University of Nebraska - Lincoln, Lincoln, NE, Dorian Garrick, Massey University, Ruakura, New Zealand, Rohan L. Fernando, Department of Animal Science, Iowa State University, Ames, IA and Hao Cheng, Department of Animal Science, University of California, Davis, Davis, CA

Linkage disequilibrium (LD), often expressed in terms of the squared correlation (r^2) between allelic values at two loci, is an important concept in many branches of genetics and genomics. Genetic drift and recombination have opposite effects on LD, and thus r^2 will keep changing until the effects of these two forces are counterbalanced. Several approximations have been used to determine the expected value of r^2 at equilibrium in the presence or absence of mutation. In this paper, we propose a probability-based approach to compute the exact distribution of allele frequencies at two loci in a finite population at any generation t conditional on the distribution at generation $t-1$. As r^2 is a function of this distribution of allele frequencies, this approach can be used to examine the distribution of r^2 over generations as it approaches equilibrium. The exact distribution of LD from our method is used to describe, quantify and compare LD at different equilibria, including equilibrium in the absence or presence of mutation, selection, and filtering by minor allele frequency. We also propose a deterministic formula for expected LD in the presence of mutation at equilibrium based on the exact distribution of LD.

PE0184: Methods: Bioinformatics

Genotyping-by-Sequencing of Highly Duplicated Genomes

Wittney D. Mays, University of Illinois, Dept of Crop Sciences, Urbana, IL, Lindsay V. Clark, Dept of Crop Sciences, University of Illinois, Urbana, IL and Erik J. Sacks, Dept of Crop Sciences, University of Illinois at Urbana-Champaign, Urbana, IL

The development of genotyping-by-sequencing (GBS) methods has facilitated genomics studies in non-model species, including polyploids. Variant and genotype calling methods have been established for autopolyploids, but remain a challenge for both recent and ancient allopolyploids (e.g. wheat, maize, soybean, Miscanthus), particularly where the reference genome contains highly similar paralogous sequences that do not pair at meiosis. Alignment of sequence tags to the appropriate position within highly duplicated reference genomes remains a challenge inadequately addressed by existing alignment software. Although some variant calling pipelines can discriminate a paralogous locus from a Mendelian locus, the detection of these paralogous loci is typically for the purpose of the exclusion of these loci from the downstream analysis of genomic studies. We explore the significance of eliminating paralogous loci in downstream analysis using a newly developed pipeline developed to sort sequence tags to their correct alignment locations based on the novel Hind/HE statistic. Through the implementation of this statistic, we present preliminary findings. The goal of this study is to evaluate the sorting pipeline's ability to properly align paralogous loci to the correct position with respect to the reference genome. We anticipate that our pipeline will result in improved genotyping quality, resulting in improved power for GWAS, GS, trait mapping, and population genetics.

PO0185: Methods: Bioinformatics

Optimising Bang for Buck in Plant Breeding

David Rodgers, Emma Mace and David Jordan, The University of Queensland, St Lucia, QLD, Australia

Plant breeding is a complex multi-stage process which involves using resources to generate new varieties that are superior to existing varieties. The breeding process involves using quantitative genetics principles to maximise

genetic gain within resource and time constraints. Thus plant breeders are faced with a complex optimisation problem with multiple solutions, the best balance between rate of genetic gain and genetic gain per dollar is not necessarily obvious and will change with varying price of inputs, scale of the program, changes in breeding targets and the advent of new technologies. Making good decisions requires good decision support tools to enable breeders to evaluate alternative scenarios both from the point of view of genetic gain and cost.

We have developed a simple tool (“Breeding Program Costing Tool”) to help breeders generate financial models of alternative breeding programs which can be used to estimate costs of running their breeding pipelines, identify high cost activities or items and provides the ability to compare different “what if” scenarios, for example, replacing single seed descent with double haploids. Once a financial model has been developed, the tool enables the breeder to duplicate elements of the model and modify these to enable comparisons of alternative scenarios to be made. The software generates a range of reports which can be used by the breeder to determine resource requirements and costs. In addition it has an interactive function that enables the cost implications of changing population size, replication levels and number of sites.

The costing tool is being implemented in multiple breeding programs in Australia and Ethiopia.

The software is freely available via <https://aussorgm.org.au/downloads/breeding-costing-tool/>

PE0186: Methods: Bioinformatics

Advancing the Game: Next-Generation Seed Company Revolutionizing Plant Breeding Readies for Expansion with Streamlined and Optimized Operations

Robert Zeigler, Ph.D., L7 Informatics Inc., Austin, TX

This poster will provide a visual and textual representation of how L7 Informatics, the leader in life + crop science informatics platforms, streamline and optimize Inari’s operations to integrate global research data, processes, and bioinformatics with L7’s scientific information management and automation software platform, Enterprise Science Platform (“ESP”).

The poster will demonstrate how Inari/L7 Informatics was able to build custom workflows utilizing L7’s scientific information management and automation software platform, Enterprise Science Platform (ESP). The poster will explain how ESP’s highly flexible and configurable process engine supports standardizing processes globally while providing flexibility to local teams to build custom workflows without the need for any programming. The poster will fully show the need, implementation process, and ongoing results of the project, such as improved data entry automation, provenance, and reporting for Inari’s high throughput lab operations. The poster will also highlight the communication facilities provided between Inari’s lab and bioinformatics teams. The poster will include various graphic images regarding workflow chains, defined reports (such as GH Transfer Manifests, Custom Sample Reports, Team Management Reports, Orphan Sample Reports, Pedigree Reports, Project Manager Reports, and KPI Reports), machine integrations, and the ease of training of key personnel. In conclusion, the poster will show with results how L7’s ESP easily supports Inari’s rapidly growing global business operations and end-to-end research processes: from early-stage R&D; through development.

PO0187: Methods: Bioinformatics

Germinate - Presenting Experimental Data from the Crop Trust Crop Wild Relatives Project

Paul Shaw¹, Sebastian Raubach¹, Benjamin Kilian², Hannes Dempewolf³, Iain Milne¹ and Gordon Stephen¹, (1)The James Hutton Institute, Dundee, United Kingdom, (2)The Global Crop Diversity Trust (GDCT), Bonn, Germany, (3)Global Crop Diversity Trust, Bonn, Germany

The Crop Wild Relatives (CWR) Project coordinated by the Crop Trust (<https://www.cwrdiversity.org/>) and supported by the Government of Norway is managing pre-breeding projects for 19 crops. These projects aim to bring back useful traits from wild relatives into their domesticated counterparts. Through thousands of crosses and backcrosses between domesticated species and their wild relatives, the project partners are generating huge volumes of characterisation, evaluation, genotypic and phenotypic data which needs to be made easily available to plant breeders and scientists. This data is collected across several countries and under various environmental conditions. Collecting and managing data is difficult. Analysing the data can pose an even greater challenge. This, however, must be addressed if pre-breeding is going to contribute to the development of sturdier crops which are better adapted for changing environments. Germinate allows these pre-breeding projects to present their experimental data in a common platform which is continually evolving with both data and the features that it makes available to users. We are developing a community of Germinate databases which contribute to this global initiative. The use of the Germinate platform means that this data will be available quickly and in meaningful ways for plant breeders and scientists. We are now developing Germinate instances in alfalfa, barley, chickpea, cowpea eggplant, finger millet, grass pea, lentil, pearl millet, pigeon pea, rice, sorghum, sunflower and wheat and will present our Germinate infrastructure which holds pre-breeding data for these species. <https://ics.hutton.ac.uk/get-germinate>

PE0188: Methods: Bioinformatics

Digitalization of Genomics Accelerates Breeding and Crop Protection Programs

Tim Zeppenfeld¹, Pierrick Gautier², Benjamin J. Adamczyk³, Florent Villiers², Michael Edmund Beck⁴ and Thomas Hartsch¹, (1)Genedata AG, Basel, Switzerland, (2)Bayer Cropscience France, Lyon, France, (3)Genedata, Lexington, MA, (4)Bayer AG, Monheim, Germany

Increasing the efficiency of breeding and crop protection programs, as well as resource management is a key challenge to address the world-wide growing demand for food and feed. Triggered by significant cost reductions in next generation sequencing (NGS), gene editing and genomic-based studies have become central to the agricultural industry. New advances in comparative genomics and cross-omics however raise the need for data integration and accessibility, to drive the development of innovative solutions based on molecular information. An enterprise software solution is required to standardize and streamline processes, integrate genotype and phenotypes data and provide easy access to the scientists and project members and provide documentation in order to be prepared for regulatory affairs. Genedata Selector® provides extensive data integration to enable comparative genomics on sequence, mutation, protein function, regulation and pathway level as well as the integration of cross-omics data processing and statistical analysis and the management of virtually unlimited number of genome sequences of plants, microbes and even microbiomes. Integrated with the bio-assay and phenotype data systems, the genotype-phenotype relationship analysis provides strategies and engineering targets for traditional and molecular plant breeding to address abiotic stress resistances and understand molecular mechanisms and develop microbiome-based methods for seed protection. Here we present Genedata Selector® and how genomics is used e.g. to support crop protection from compound identification, mode-of-action analysis towards resistance monitoring in fungicide research.

PO0189: Methods: Bioinformatics

Using a Real-Time NGS Platform for Interpretative Analysis of Complex Plant Genomes

Shawn Quinn, Curio Genomics, Dexter, MI

Plant genomes are often complex and analysis with traditional bioinformatics tools proves difficult. Leveraging decades of big data software development experience, we describe the utility of a fully scalable, real-time, NGS analysis platform for analyzing complex plant genomes. This advancement unleashes new opportunities for researchers seeking to address critical global challenges to develop more productive and resilient crops in a time of climate change and further population growth.

Among the plant genomes, none are more challenging to analyze than the wheat (*Triticum aestivum* L.) genome due to its large size, 85% repeating sequences, and polyploidy nature. The publication of the wheat genome reference and related annotations by the International Wheat Genome Sequencing Consortium (IWGSC) has made the analysis of the wheat genome feasible, in theory.

Here we describe how we overcame the challenges of read mapping (both for DNA-Seq and RNA-Seq libraries) and read alignment visualization when dealing with the "large chromosome" complexity of the wheat genome. Additionally, we show a novel approach to variant calling, coverage analysis, and gene expression calculation in hexaploid species. Including the dynamic incorporation of the IWGSC reference and annotation sets, we share several research examples developed using the [Curio Genomics](#) platform. Based on collaboration with IWGSC members, we highlight powerful interpretive results and data visualizations, including an approach for filtering variants by predicted biological consequences.

PE0190: Methods: Bioinformatics

Development of Plant Genome Portal Site PlantGARDEN

Daijiro Harada¹, Hisako Ichihara², Akihiro Nakaya², Andrea Ghelfi¹, Hideki Hirakawa¹, Satoshi Tabata¹ and Sachiko Isobe¹, (1)Kazusa DNA Research Institute, Kisarazu, Chiba, Japan, (2)Graduate School of Medicine, Osaka University, Suita, Osaka, Japan

In recent years, genome analysis can be performed more inexpensively and efficiently with the development of sequencing technologies. *De novo* whole genome assembly has become common in non-model species. Currently (April 2019), more than 380 plant genomes have been deciphered, and it is considered that the numbers of assembled genomes will keep increasing in future. In addition, resequencing is commonly performed in plant species which reference genomes have already been constructed. Not only quantity, the accuracy of genome sequences is also improved year by year.

The explosive increase of genome sequences data requires portal websites that completely covers the plant genomes. There are a few portal websites available for plant genomes such as Phytozome (<https://phytozome.jgi.doe.gov/pz/portal.html>) and Ensembl (<https://asia.ensembl.org/index.html>). However, these are more for experts and sometimes difficult to beginners. Therefore we are developing a non-expert friendly plant genome integration database, Plant GARDEN (<https://plantgarden.jp/en/>). The database includes assembled genomes, genes, DNA markers and trait related loci information. Gene sequence similarities across plant species is also available in the DB. The English beta version is opened in June 2019 for nine plant species. Currently, only limited information is available, however, Plant GARDEN is aimed to include most of plant genome information in future. This work is supported by the JST Life Science Database Integration Project (17934006).

PO0191: Methods: Bioinformatics

The Germinate Plant Resources Data Platform

Sebastian Raubach, Gordon Stephen, Iain Milne and Paul Shaw, The James Hutton Institute, Dundee, United Kingdom

Germinate is an open source plant database infrastructure and application platform on which complex data from genetic resources collections can be stored, queried, and visualized using common, reusable, programming components. Germinate utilizes modern web and database standards to provide a standard architecture and high-performance web-based user interface and analytics functionality across a wide variety of data types. Germinate is fully compliant with the Multi-Crop Passport Descriptors (MCPD V.2.1) developed by the Food and Agriculture Organization (FAO). The single point of access for all background data on genetic resources collections that Germinate implements provides a valuable digital curation resource as well as offering essential continuity on how new data is recorded and archived.

Germinate not only acts as storage for experimental data but also offers connections to visual analytics tools such as Flapjack for graphical genotyping and Helium for pedigree visualization. We have prioritized the development of tools to allow the export of data in formats suitable for analysis in statistical analysis tools such as R.

Germinate has been developed as an extensible platform, with an aim of offering an off-the-shelf solution for the storage of genetic resources data and has been used in several projects across a wide variety of species including barley, wheat, potato, maize, eggplant and sunflower.

PE0192: Methods: Bioinformatics

GnpIS, a Multispecies Integrative Information System for Plant and Fungi

Mélanie Buy¹, Emilie Millet², François Tardieu³, Claude Welcker⁴, Cyril Pommier⁵ and Anne-Francoise Adam-Blondon⁵, (1)URGI, INRA, Versailles, France, (2)WUR, WAGENINGEN, Netherlands, (3)INRA, Montpellier, France, (4)INRA, montpellier, France, (5)URGI, INRA, Université Paris-Saclay, Versailles, France

[GnpIS](#) is a multispecies integrative information system which allows plant biologists to share genomic, genetic and phenotypic complete datasets. Final phenotypic datasets from the [DROPS European project](#) have been therefore published in GnpIS using the [MIAPPE](#) (Minimal Information About Plant Phenotyping Experiment) format. It encompasses 19 trials in Europe and Chile which survey 246 maize hybrids. Indeed, 24 variables are measured from these genotypes, both phenotypic including morphological, phenological and yield traits, together with environmental variables including light, growing temperature and soil conductance. Two watering conditions was implemented: irrigated and rainfed treatments.

The data is displayed by level (microplot, genotype) with one level by matrix that integrates the data from multiple trial and years. These matrix show results for all accessions within a trial, by treatment (rainfed or watered parcels for DROPS data), for all observed variables. The data can be refined in multiple ways, for example, by phenotyping campaign year. Information about trial can be consulted on a dedicated webpage, including site, description, genotypes and contact sections. Some open format additional data files can be also downloaded and an export in MIAPPE/ISA-TAB or tabular format is possible for each matrix. Each germplasm is linked to a webpage reporting DOI, holding institution and taxonomy. A widget allows to navigate among variables and to browse information in BrAPI format (trait, variable, scale) and cross-reference leading to Crop Ontology variables are provided. Finally these phenotypic data are linked to maize genotyping and GWAS data already publicly available in GnpIS.

PO0193: Methods: Bioinformatics

Genome-Wide Identification of Candidate Genes with Resistance to Biotrophic and Necrotrophic Pathogens in Wheat

Rong-Cai Yang, University of Alberta, Edmonton, AB, Canada

There are two groups of plant pathogens depending on their lifestyles or modes of nutrition acquisition. The first group, obligate biotrophs, develop coevolving pathogenesis mechanisms that interfere host physiology while maintaining host viability. The second group, necrotrophs, kill host cells prior to or during colonization and absorb nutrients from dead host tissues and they often produce phytotoxins and cell wall-degrading enzymes (CWDEs), leading to the host cell death and nutrient leakage. In response to pathogen attack, plant resistance (R) proteins recognize specific pathogen proteins (effectors) and elicit an effector-triggered immunity (ETI) against many biotrophic and hemibiotrophic pathogens, whereas the necrotrophic diseases arise from the recognition of necrotrophic effectors (NEs) [also known as host-specific toxins (HSTs)] by dominant host sensitivity genes. This presentation reports a genome-wide search for candidate genes with resistance to three biotrophic rust pathogens, *Puccinia graminis* f. sp. *tritici* (stem rust), *P. triticina* (leaf rust) and *P. striiformis* f. sp. *tritici* (stripe or yellow rust), and two necrotrophic pathogens, *Pyrenophora tritici-repentis* (Died.) Drechs. (tan spot) and *Fusarium graminearum* (Fusarium head blight, FHB) in common wheat (*Triticum aestivum* L.). These are among the most damaging crop diseases in western Canada and around the world. The candidate genes identified for the biotrophic vs. necrotrophic pathogens will be compared and contrasted in terms of their genome-wide distributions and gene families. This assessment will allow for molecular and functional insights into the possible co-occurrence of the genes with resistance to biotrophs and necrotrophs within the same cultivars.

PE0194: Methods: Bioinformatics

Reference Genome Selection Method for Accurate Pathogen Typing through Whole-Genome Sequencing Data

Sojin Ahn¹, So Yun Jhang¹ and Heebal Kim^{1,2}, (1)Interdisciplinary Program in Bioinformatics, Seoul National University, Seoul, South Korea, (2)Department of Agricultural Biotechnology and Research Institute of Agriculture and Life Sciences, Seoul National University, Seoul, South Korea

Over the last years, rapid technology development has provided whole-genome sequencing (WGS) as a high-resolution subtyping and real-time surveillance of foodborne pathogens. In more detail, the single nucleotide polymorphism (SNP)-based subtyping using WGS that compares sequences to the reference genome has become popular. However, unlike other organisms, bacteria shares genes between different species through horizontal gene transfer (HGT). Thus, the genomic information can be varied and affect the results of pathogen typing using SNP depending on which reference genome is selected. Therefore, it is important to choose the appropriate reference genome sequence for an accurate pathogen typing.

In this regard, we updated our previously created pathogen typing program called SNP identification for Strain Typing (SNPing) with an algorithm of selecting the reference genome based on the mapping rates. A total of 2,819 genome sequences of the foodborne pathogens were used as the reference genome candidate and compared to new WGS samples. During the process of selecting a proper reference genome, the highest mapping rate of the newly mapped sequencing read data was chosen. In order to validate this method, we created and tested simulation data using *Clostridium perfringens* genome data sets, which contain 100 SNPs and INDELs. As a result, all the simulated genomes were selected to the original data correctly, with over 99% mapping rate.

Base on this result, we have improved the previous web-based pipeline program with the supplementing an algorithm of selecting the appropriate reference genome. We believe that our additional method of selecting a reference genome based on the mapping rate will be a beneficial analytical tool for pathogen typing. This research was supported by a grant 19162MFDS037 from Ministry of Food and Drug Safety in 2019.

PO0195: Methods: Bioinformatics

Impact of Genotype Calling Methodologies on Genomic Selection and GWAS Sensitivity in Polyploids

Joyce Njuguna¹, Lindsay V. Clark¹, Xiaoli Jin², Karen Koefoed Petersen³, Kossonou Anzoua⁴, Pavel Chebukin⁵, Elena Dzyubenko⁵, Nicolay Dzyubenko⁵, Bimal K. Ghimire⁶, Kweon Heo⁷, Douglas A. Johnson⁸, Uffe Jørgensen⁹, Jens B. Kjeldsen⁹, Hironori Nagano¹⁰, Andrey Sabitov⁵, Toshihiko Yamada¹¹, Ji Hye Yoo¹², Chang Yeon Yu¹² and Erik J. Sacks¹³, (1)Dept of Crop Sciences, University of Illinois, Urbana, IL, (2)Zhejiang University, Hangzhou, China, (3)Aarhus University, Årsløv, Denmark, (4)Hokkaido University, Sapporo, Hokkaido, Japan, (5)Vavilov All-Russian Institute of Plant Genetic Resources, St. Petersburg, Russian Federation, (6)Konkuk University, Gwangjin-gu, South Korea, (7)Kangwon National University, Chuncheon, South Korea, (8)Utah State University, Logan, UT, (9)Aarhus University, Tjele, Denmark, (10)Hokkaido University, Hokkido, Japan, (11)Field Science Center for Northern Biosphere, Hokkaido University, Sapporo, Japan, (12)Kangwon National University, Chuncheon, Korea, Republic of (South), (13)Dept of Crop Sciences, University of Illinois at Urbana-Champaign, Urbana, IL

Discovery and analysis of the genetic variants underlying agriculturally important traits is key to molecular breeding and improvement of polyploid crops. Reduced representation approaches have provided fast and cost-efficient genotyping using next generation sequencing (NGS). However, accurate genotype calling from NGS data is challenging, particularly in polyploid species where there is uncertainty in allele copy number in heterozygotes due to the genome multiplicity, complex inheritance patterns and low sequencing depth. Recently developed Bayesian statistical methods estimate genotype likelihoods, while incorporating population level data to obtain measures of uncertainty associated with genotypes. In this study we focus on three Bayesian algorithms, implemented in polyRAD, EBG and updog respectively, and demonstrate the impact of their allele dosage estimation on genome wide association analysis (GWAS) and genomic selection (GS). We compare the performance of these Bayesian genotype callers to naive genotype calling approaches similar to those implemented in the Genome Analysis Toolkit

(GATK). The present study also evaluates the use of posterior mean probabilities (continuous) as opposed to discrete genotypes to increase power for GS and GWAS. We implement the genotype calling on a dataset from a diversity panel of *M. sacchariflorus* and perform association analysis of several simulated traits. This study demonstrates that the use of posterior mean probabilities increases power for GS, hence facilitating the acceleration of crop breeding.

PE0196: Methods: Bioinformatics

Using Long Reads to SNP Calling and Haplotype Reconstruction in a Tetraploid Potato Cultivar

Fernanda A. de F. Guedes, Richard G.F. Visser and Richard Finkers, Wageningen University & Research, Plant Breeding, Wageningen, Netherlands

The coming of third generation sequence technologies, like Nanopore platform, have been contributing to improve genome assemblies, generating longer scaffolds and allowing variant discovery. On the other side, the higher error rate and number of small indels of Nanopore long reads make them more intractable to SNP calling. Using these long reads for SNP calling and SNP-based haplotyping of polyploid genomes can be even more challenging, specially because of the false positive SNPs. The goal of this work is to use Nanopore long reads to call a minimal subset of biallelic SNPs to be used to reconstruct the haplotypes for a autotetraploid potato cultivar. First, tetraploid potato long reads were mapped to DM potato genome using Graphmap. Based on summarized read counts for each position, only the putative biallelic (REF:0.75, ALT:0.25 read count ratio) SNPs whose depth was above the user-defined threshold were selected. Then, each chromosome was split into 10 kb windows. Consecutive SNPs of each window were combined in triplets and used to query aligned Nanopore reads. The classification of triplets count profiles using the self-organizing maps method allowed to identify genomic regions with four haplotypes and select the SNPs for haplotype reconstruction, the final step carried out with wtdg2 tool. The developed approach is focused on the genomic regions with four different haplotypes and was applied in a 80-kb test region of chr01 of DM, allowing the reconstruction of expected haplotypes for six out of eight regions and the merging of consecutive regions, using four SNPs per region. To enlarge the application of the approach to the chromosome-scale, a C implementation is in progress, including user-defined parameters like the window size and depth threshold.

PO0197: Methods: Bioinformatics

Plant Omics Databases: Plant Omics Data Center (PODC), CATchUP and TOMATOMICS

Toshiharu Endo¹, Misa Saito¹, Maasa Kanno¹, Shizuka Koshimizu¹, Kong Bihe¹, Shun Ohki¹, Eiji Nambara², Hajime Ohyanagi^{1,3}, Koh Aoki⁴, Hiroshi Ezura⁵ and **Kentaro Yano¹**, (1)Sch. of Agri., Meiji Univ., Kawasaki, Japan, (2)Dept. of Cell & Systems Biology, University of Toronto, Toronto, ON, Canada, (3)Computational Bioscience Research Center, King Abdullah University of Science and Technology, Thuwal, Saudi Arabia, (4)Grad. Sch. of Life and Environ. Sci., Osaka Pref. Univ., Sakai, Japan, (5)Faculty of Life and Environmental Sciences, University of Tsukuba, Tsukuba, Japan

Here, we introduce three databases ‘Plant Omics Data Center’ (PODC, <http://bioinf.mind.meiji.ac.jp/podc/>), CATchUP (<http://plantomics.mind.meiji.ac.jp/CATchUP/>) and TOMATOMICS (<http://bioinf.mind.meiji.ac.jp/tomatomics/>). PODC contains both the knowledge-based functional annotations and omics information on 13 plant species. To provide gene expression profiles, RNA-Seq datasets for the species were collected from the Sequence Read Archive (SRA), and were categorized by Plant Ontology and Plant Environmental Ontology according to the information of mRNA samples examined (cultivar, developmental stage, sampling time, organ/tissue, treatment). With the gene expression profiles obtained from the RNA-Seq data, gene expression networks (GENs) in each species were constructed based on the similarity of expression profiles between genes. We also assigned knowledge-based functional annotations including the information of transcription factors and cis-elements of genes by natural language processing techniques and manual curation of literature. The transcription factors are also able to be integrated in GENs. By using the information of orthologues among species, GENs from different species can be connected and compared for identification of evolutionary conservation in the similarity of the sub-networks. A database CATchUP provides the information of spatiotemporally expressed genes in eight plant species. With a developed statistical method, approximately 70,000 transcripts were extracted as candidates of spatiotemporally expressed genes. The main information in the tomato database TOMATOMICS documents experimental resources (cDNA clones, which are provided from The National BioResource Project (NBRP) in

Japan, generated from the cultivar Micro-Tom, and gene structural annotations (TMCS v. 1.2.1) for the reference genome (the cultivar Heinz 1706). By accessing the TOMATOMA database, users can easily identify tomato experimental resources and predicted gene structures.

This research was supported by KAKENHI (19H04870) and the Research Funding for Computational Software from Meiji University.

PE0198: Methods: Bioinformatics

Time-Series Association Modeling and Exa-Scale Environmental Clustering to Create Designer Idiotypes in North American *Populus trichocarpa*

Jared Streich¹, Gabriel Gazolla¹, Ben Garcia¹, Annie Furches², David Kainer¹, Ashley Cliff², Jonathan Romero², Micheal Garvin³, Sharlee Climer⁴, Peter Thornton¹ and Daniel Jacobson¹, (1)Oak Ridge National Laboratory, Oak Ridge, TN, (2)University of Tennessee/Oak Ridge National Laboratory, Oak Ridge, TN, (3)Oak Ridge National Laboratory, Oak Ridge National Laboratory, TN, (4)University of Missouri St. Louis, St. Louis, MO

We developed a method to create designer climate idiotypes for specific geography through the integration of daily resolution time-series genotype by environment association models and global climate clustering. We created Genome-Wide Association Time-series Studies (GWATS) and Periodic Genotype by Environment Association (PGEA) to analyze daily resolution time-series data of climate variables to find locally adaptive candidate alleles including the time of year they convey potential adaptiveness. Our time-series association method greatly reduces complexity of association output through wave-form analytics to parse candidates into subgroups based on -log transform shape values, in which we found many sets of adaptive candidates for several environmental stresses. Coupling this method with an exa-scale DUO calculation at 2.31×10^{18} operations/second with a final output of 2.1×10^{22} operations on the Summit supercomputer at Oak Ridge National Laboratory to calculate similarity values of all dry land. DUO is a long-vector similarity calculation metric in the CoMET software that implements rapid small matrix pair-wise GPU based calculations. We encoded every global square kilometer into 414,640 long vectors composed of daily resolution from 71 climate variables. We then used MCL clustering to generate several climate cluster maps with 69, 133, 340, and 717 different suites of global climate zones for all ≈ 156 million km². With suites of adaptive alleles for several environmental stresses we then examined specific geographic areas and ran genomic prediction models to create lists of breeding parents for targeted geographic regions.

PO0199: Methods: Bioinformatics

Identifying Drivers of Latitudinally Varying Traits in Poplar

Cai W. John, University of Tennessee/Oak Ridge National Labs, Oak Ridge, TN

Poplar (*Populus trichocarpa*) phenotypic traits have been shown to vary along a North-South gradient. These traits that vary along a North-South gradient are economically pertinent as it allows manifesting longer grow cycles and photosynthetic periods in Northern genotypes and potentially enhancing these traits Southern genotypes. To determine a potential genetic basis, we analyzed single nucleotide polymorphism (SNP) data and expression data from 917 poplar field samples derived from across the Northwest U.S. and Canada. Population structure of the SNP data revealed two clear subpopulations that correspond to the river system the field specimens were collected from. We were also able to predict the latitudinal group of each sample with 93% confidence using RNA-seq data derived from leaf samples, indicating additional transcriptional differences that correspond to sample source location. Following up on both signals, GWAS was performed using GEMMA to identify 77 SNPs significantly associated ($p\text{-value} < 6.02 \times 10^{-9}$) with latitude. Moving forward, we will annotate these SNPs to identify their functional role and use SNPs to elucidate the directionality of the change. An eQTN analysis will then be used to determine whether these significant SNPs correspond to either observed transcriptional differences or our previous GWAS results using these samples, which were performed for a number of important traits, to determine what traits these significant SNPs are associated with.

PE0200: Methods: Bioinformatics

Computational Profiling and Characterization of MicroRNAs and their Targets in Lettuce (*Lactuca sativa* L.)

Dr. Muhammad Din, Government Postgraduate Science College, Quetta, Quetta, Pakistan

MicroRNAs (miRNAs) are tiny, endogenous, non-protein-coding, and functionally negative regulators of posttranscriptional gene regulation. Lettuce (*Lactuca sativa* L.) is one of the most popular vegetable worldwide. Though Lettuce is an important dietary food, its miRNAs and their targets have not been well studied. Here we report a comparative genomics approach that was applied to explore miRNAs and their targeted proteins in Lettuce. A total of 60 new miRNAs from 45 families were identified and characterized from the expressed sequence tags. All 60 miRNAs were observed, along with stable stem-loop precursor structures, whose lengths ranged from 54 to 470 nt with an average of 142 nt. Mature miRNAs lengths ranged from 18 to 26 nt with an average of 20 nt. Later, a total of 235 potential targets were predicted for these new 60 Lettuce miRNAs. These targets were involved in regulation, metabolism, transcription factors, growth and development, and other physiological processes. These miRNAs and their targets will be useful to fine-tune Lettuce for better food content and higher nutritional value as well as biotic and abiotic stress resistance.

PO0201: Methods: Bioinformatics

PacBio Single Molecule Real-Time Sequencing of the Sweet Orange Leaf Transcriptome

Saroj Parajuli¹, Ze Peng¹, Fred G. Gmitter² and Zhanao Deng¹, (1)University of Florida, Wimauma, FL, (2)University of Florida, IFAS-CREC, Lake Alfred, FL

PacBio single molecule real-time sequencing has become a powerful tool to obtain a comprehensive view of transcriptome profiles, including full-length transcripts, alternative splicing events, transcriptional start sites, transcriptional termination sites, and poly-adenylation sites. To our best knowledge, application of this technology has not been reported in *Citrus*, a genus containing many important fruit crops for fresh fruit and juice production. We applied this technology to *Citrus* and sequenced the leaf transcriptome of sweet orange (*C. sinensis*). A total of 201,782 full length non-chimeric reads were generated with an average read length of 2.7 kb. Further processing yielded 18,919 high-quality and 87,648 low quality consensus isoforms. By mapping the high-quality isoforms to the *C. sinensis* reference genome (JGI V1.1), we found that most isoforms are full-splice-match (58.3%), followed by novel isoforms (33.1%). Most sweet orange gene models (73.1%) had a single isoform, 17.3% had two isoforms, and the remaining gene models had more than two isoforms. Further investigation of full-splice match isoforms showed that for the majority of isoforms, both the transcriptional start sites and transcriptional termination sites are upstream of their annotated positions in the genome, implying the need to improve the current genome annotations. Moreover, we identified 324 novel genes (not annotated in the current genome), 55 fusion transcripts, 296 transcription factors, and 3,285 long non-coding RNA. Our research provided insights into the sweet orange leaf transcriptome and guidance for further improvement of the sweet orange genome annotation and future isoform sequencing studies in *Citrus*.

PE0202: Methods: Bioinformatics

Accurate Characterization of Expression and Alternative Splicing in Arabidopsis for Protein Coding and Long Non-Coding RNAs

Runxuan Zhang¹, Wenbin Guo², Nikoleta Tzioutziou², Juan Carlos Entizne², Cristiane Calixto², Allan James³, Csaba Hornyik¹, Hugh Nimmo³ and John Brown², (1)The James Hutton Institute, Dundee, United Kingdom, (2)University of Dundee, Dundee, United Kingdom, (3)University of Glasgow, Glasgow, United Kingdom

Understanding the current limitations of RNA-seq is crucial for reliable analysis. We have developed several computational resources, methods and tools to address the challenges of RNA-seq data analysis, with the emphasis on plant species. We have carried out:

- A BBSRC funded project to construct an automated pipeline with multiple assemblers to capture the diversity of transcripts from different sources and technologies and stringent filters to construct a comprehensive Reference Transcript Dataset (RTD) for plants. Extensive experimental validation showed

that RTDs constructed using our method outperform other available transcriptomes in RNA-seq analysis in quantification accuracy [1][2].

- A cutting-edge pipeline (*3D RNA-seq*) [3] for differential gene expression and alternative splicing analysis. *3D RNA-seq* incorporates the state-of-the-art methodologies while remaining simple and rapid. It allows (lab) biologists with no programming skills to perform a complete differential expression analysis of RNA-seq data in 3 days.

These tools/methods enabled the discovery of massive and rapid expression and AS responses to cold in *Arabidopsis* and identification of hundreds of genes with very early changes in expression/AS, including numerous novel cold-responsive transcription factors and splicing factors/RNA binding protein genes [4]. We also demonstrated cold induced changes in expression as well as AS of pri-miRNAs and lncRNAs [5].

[1] Zhang et al. (2017) *Nucleic Acid Research*, 45 (9): 5061-5073; [2] Flores et al (2019) *BioRxiv* <https://doi.org/10.1101/638106>; [3] Guo et al. (2019) *BioRxiv*. <https://doi.org/10.1101/656686>; [4] Calixto et al. (2018) *Plant Cell*, 30(7):1424-1444; [5] Calixto et al. (2019) *Front. Plant Sci.* 10:235

PO0203: Methods: Bioinformatics

Genomic Prediction and Variance Component Estimation Using Inbred Lines Without Heterozygous Genotypes

Yang Da, Department of Animal Science, University of Minnesota, Saint Paul, MN

Current methods for genomic prediction and estimation are developed for general populations with homozygous and heterozygous SNP genotypes, but self-fertilization plant species generally do not have heterozygous SNP genotypes. This raises question whether the methods for genomic prediction and variance component estimation for populations with heterozygous genotypes also apply to populations without heterozygous genotypes. This study conducted a theoretical analysis of two existing methods (Methods 1 and 2) and one new method (Method 3) for genomic prediction and estimation in populations without heterozygous genotypes. Method 1 uses the total additive variance in a random population under Hardy-Weinberg equilibrium and without inbreeding as the denominator of genomic additive relationship matrix, and Method 2 uses the average of the diagonal elements of WW' as the denominator, where W is the model matrix of SNP additive effects from quantitative genetics partition. Method 3 is a new method and uses genomic relationship matrix specific to populations without heterozygous genotypes. The results showed all three methods yield identical GBLUP and reliability values, and Methods 2 and 3 also have identical genomic relationships, additive variance and heritability for populations without heterozygous genotypes. For Method 1, the diagonal elements of the relationship matrix are twice as large and the additive variance is half as large as those of Methods 2 and 3. Consequently, heritability estimate of Method 1 is smaller than that of Methods 2 and 3. The heritability estimate of Method 1 can be converted to that of Methods 2 and 3 or vice versa. The formulae for the conversions are: $h_1 = h_f/(2-h_f)$, and $h_f = 2(h_1)/(1+h_1)$, where h_1 = additive heritability from Method 1, and h_f = additive heritability from Method 2 or 3. Variance components and heritability estimates of Method 1 are those for a random population even though the data is from an inbred population without heterozygous genotypes. Methods 2 and 3 yield estimates of additive variance and heritability for populations without heterozygous genotypes but its genomic relationship matrix is a genomic coancestry matrix rather than a genomic additive relationship matrix. With correct interpretation and conversion of genomic relationships and estimates of additive variance and heritability, either Method 1 or 2, as implemented by GVCBLUP that is freely available at <https://animalgene.umn.edu/gvcclub>, can provide correct analysis of genomic prediction and estimation of variance components and heritability for populations without heterozygous genotypes. Method 3 for populations without heterozygous genotypes does not need to be implemented.

PE0204: Methods: Bioinformatics

Detection of Interspecies-Level Directional Selection Based on Single Nucleotide Polymorphism and Divergence from the Reconstructed Ancestral Sequence

DongAhn Yoo, Chul Lee and Heebal Kim, Seoul National University, Seoul, South Korea

Triggered by the large-scale international consortiums including vertebrate genome project (VGP) generating high quality genome assemblies, bottleneck of a scientific research is now shifting to bioinformatic analysis. Simple and effective comparative genomics analysis is needed. McDonald and Kreitman test (MKT) in this sense is a great tool for detecting positive selection. Here, we developed a new MKT pipeline that can boost up its performance via ancestral sequence reconstruction. The pipeline was assessed by investigating recent human evolution using forward simulation data and population genome data of four Great Ape species (Human, Chimpanzee, Bonobo and Gorilla). In the simulation data, we traced the accumulated variants to compare performance of the MKTs in detecting true positive selection. We observed better performance of novel MKT compared to the standard approach. In the public data, we detected positive selection in 900 genes and 700 genes for human with novel and standard approach, respectively. Based on sequence conservation of phylogenetic tree of great ape lineage, we found that approximately half of these functional substitutions are likely to emerge outside human lineage and result in false positive calls of up to ~300 genes in standard approach. Similarly, we found non-functional polymorphisms could also influence up to 900 non-selected genes in standard test including *SALL3* related to GO term “neurogenesis”. In comparison with maximum likelihood-based dN/dS, higher number of genes were commonly found with the novel approach compared to the standard MKT. The result from this preliminary analysis highlights the robustness and strength of the novel pipeline over the standard MKT.

PO0205: Methods: Bioinformatics

Test of the Effect of Recombination and Migration on Genomic Islands of Differentiation Under Divergent Selection Using Experimental Evolution in Yeast (*Saccharomyces cerevisiae*)

Ousseini Issaka Salia¹, Martyna Lukaszewicz¹ and University of Idaho Yeast Project Authors,
(1)University of Idaho, Moscow, ID

Genomic data provides the possibility to learn how environmental conditions structure the genome of organisms. For example, markers exhibiting elevated genetic differentiation between populations have traditionally been used to detect loci under divergent selection. The effects of this divergent selection can be observed over large, physically linked regions on chromosomes, a phenomenon known as genomic islands of differentiation. Both theory and empirical studies support the view that genomic islands are the result of complex interactions among selection, migration, and recombination. First, we tested whether genomic islands of differentiation can form from standing genetic variation in an admixed population. And second, we tested the effect of migration and recombination on genomic islands of differentiation.

To test these hypotheses, we crossed two haploid strains of budding yeast to generate an admixed polymorphic population from their F2 offspring. We then evolved replicate populations in two different stress environments - sodium dodecyl sulfate (SDS) and sodium chloride (NaCl)- for 12 days and looked for genomic regions that differentially responded to these environments. At the same time, we used three sources of genomic data on recombination rate heterogeneity across the genome for this specific yeast cross, and tested for a relationship between local recombination rate and genomics islands of differentiation. We quantified genomic islands of differentiation across the genome, evolving from standing variation under divergent selection, without varying levels of sexual reproduction and migration between divergent environments. Genomic islands were consistent across replicates in the absence of migration or sexual reproduction, but large-scale variation in genomic island distribution was observed with these factors added. Location of genomic islands was consistent with annotated genes associated with osmotic and cell wall stress. As predicted, we found that high recombination and migration rates decrease the size of genomic islands of differentiation, although these effects were not consistent across all populations. Overall we found that evolution is highly variable and stochastic at the genomic level, even when starting from a common ancestor and using adaptation from standing variation.

PE0206: Methods: Bioinformatics

Apollo Provides Collaborative Genome Annotation Editing with the Power of Jbrowse

Nathan A Dunn, Lawrence Berkeley National Laboratory, Berkeley, CA

JBrowse is a powerful genome viewer that allows researchers to view and share their genomes as well as customize that experience. The visualization of genomic elements is an important step to be able to more precisely describe

annotated genomes, which is vital for accurately modeling the biological function of genomic elements. Apollo (<https://github.com/GMOD/Apollo>) is a web-based genome annotation editor that utilizes JBrowse to allow users to refine their genome annotations using JBrowse tracks as evidence, such as genomic and transcriptome elements and predictive models.

In addition to the ability to visually review diverse sets of information, Apollo is also a collaborative tool with many features that improve the efficiency of an annotation project including real-time collaborative editing, the ability to promote search results directly as evidence, a revertible and visual history of genomic edits, and many automated structural editing operations. Additionally, functional annotations including Gene Ontology (GO) annotations are supported, as well as the ability to populate arbitrary metadata that can be predefined to support a research group's workflow.

Here, we will show how to get a project going quickly using Docker. This includes adding users, uploading genomes (FASTA) and evidence (GFF3, VCF, BAM) to the user-interface, creating annotations and exporting those annotations with author attributions as GFF3 for structural annotations and GPAD2/GPI2 for GO annotations. Finally, we will show how to take advantage of Apollo's web services using the python-apollo library (<https://pypi.org/project/apollo/>).

Apollo is used in hundreds of genome annotation projects around the world, ranging from the annotation of a single species to lineage-specific efforts supporting the annotation of dozens of genomes.

PO0207: Methods: Bioinformatics

Targets of Opportunity: The Birth of Orphan Genes, Exemplified in Yeast, Arabidopsis, Maize and Humans.

Eve Syrkin Wurtele¹, Urminder Singh², Priyanka Bhandary¹, Jing Li¹ and Arun S. Seetharam¹, (1)Iowa State University, Ames, IA, (2)Iowa State University

Each organism contains genes that encode proteins with no homolog in other species ("orphan genes"). Some orphan genes have arisen *de novo* from non-genic material, others from within lncRNAs, others from novel reading frames within protein-coding genes, while others result from ultra-rapid mutation of existing genes. The challenge of distinguishing protein-coding orphan genes in genomes and predicting their functions is immense, resulting in under-appreciation of their importance.

Many transcripts containing open reading frames (ORFs) that bear no homology to other proteins are expressed and translated, but are annotated as lncRNAs, or not annotated as genes at all. Under the premise that some of these are protein-coding orphan genes, we created an aggregated dataset from RNA-seq raw reads in NCBI-Sequence Read Archive from four diverse eukaryotes: *Saccharomyces cerevisiae*, *Homo sapiens*, *Arabidopsis thaliana* and *Zea mays*. These datasets, comprising between 3000 and 15,000 samples for each species, were realigned to the respective genomes, and ORFs within these transcripts were subjected to phylostratigraphic analysis. In yeast alone, 15,806 transcripts contain ORFs inferred to be orphans ("orphan-ORFs"), about 40% of which are ribosome-bound. Taken together, these data help distinguish those transcripts that may be protein-coding proto-genes or genes, from lncRNAs, and from the random transcription that may provide fodder for new genes. We provide aggregated RNA-Seq datasets with sample metadata in MetaOmGraph (MOG), a tool enabling powerful, interactive, statistical analysis and visualization of specific transcripts under user-selected conditions. This approach enables reuse of these data for exploratory discovery and provides a rich context for experimentalists to identify and make novel, experimentally-testable hypotheses about candidate genes.

PE0208: Methods: Bioinformatics

Customer Success Story for Diploid Vegetable Seed Genotyping Data Analysis at Scale

Wouter Uten, UgenTec, Inc., Cambridge, MA

Prior to launch, several leading crop genotyping companies collaborated with UgenTec during the Genotyper software open access program. These at-scale genotyping pilot projects included running the FastFinder Genotyper

algorithms against large amounts of live genotyping data sets. These initial studies were performed with multiple high-volume genotyping laboratories. With one early adopter of the platform, samples tested included multiple different diploid vegetable seeds. In this study, automated analysis of more than 1500 production 1536-well plates was compared to the status quo analysis SOP.

PO0209: Methods: Bioinformatics

GeneSieve: A Systematic, in silico Approach for Identifying Candidate Genes in Plants

Brian Nadon, Genomics and Bioinformatics Research Unit, USDA-ARS, Athens, GA and Justin N Vaughn, USDA-ARS/University of Georgia, Athens, GA

A fundamental goal of genetics is to determine which portions of an organism's genome are responsible for particular traits. Inexpensive sequencing and genotyping techniques have led to the rapid growth of identified quantitative trait loci (QTL) affecting various traits in crop plants. However, precisely identifying candidate genes that could directly affect these traits is comparatively much more difficult, often requiring time- and labor-intensive functional analyses of many different genes. A bioinformatics-based, systematic, and efficient approach to identifying which genes in a mapped QTL are likely to affect a trait of interest is thus critical for advancing crop genetics and breeding. We therefore propose GeneSieve, a webtool that leverages existing association or linkage mapping, phenotype, and expression data to identify and quantify which genes in a mapped region are most likely to be candidate genes suitable for further study. Users of GeneSieve submit the results of a mapping study they have carried out in the form of a mapped sequence and a detailed description of the phenotype measured. GeneSieve then *ab initio* annotates the sequence for predicted genes and searches for homologs of these genes contained in preexisting QTL databases. The preexisting database information stored in GeneSieve includes gene and QTL data from model species like *Arabidopsis*, soybean, or maize, and uses TF-IDF text analysis to determine whether these previously-mapped QTL in these model species are similar to the user's input sequence and phenotype information. Finally, by storing and pre-calculating coexpression matrices for these model species using large amounts of expression data from NCBI's SRA database, GeneSieve finds genes associated with the trait of interest that may not be homologous to the candidate gene but are still coexpressed to broaden and strengthen the evidence for each candidate gene target. By combining the evidence from these homology searching, phenotype text matching, and coexpression analyses, the genes within the user's input sequence can be scored and ranked for how likely they are to be candidate genes worth further investigation. GeneSieve represents a valuable tool for breeders looking to improve the strength of their markers and selection processes, or for geneticists to discover insights that can guide further study of gene function and even provide targets for gene editing.

PE0210: Microbes and Pathogens

Automated and Manual Purification of Bacterial DNA from Phytoplasma Infected Plant Tissue

Fritz M. Wells, Promega, Madison, WI

Phytoplasmas are phloem-limited bacterial plant pathogens requiring an insect vector for transmission. Such bacterial pathogens, and the diseases they cause, have the potential to severely decrease agricultural yields. Thus, it is essential to identify infected plants. Phytoplasmas are relatively understudied organisms due to their inability to be cultured in vitro. Therefore, alternative methods are required for detection and characterization of these organisms. Also, while phytoplasma infected plants can exhibit observable symptoms, such indications of disease are not always expressed or clearly visible. For these reasons, successful purification and detection of phytoplasma DNA isolated from plant tissue is a valuable tool for identification of phytoplasma infected plants and analysis of phytoplasmas in general. Here we describe three methods for successful purification of phytoplasma DNA from infected Ash tree (*Fraxinus* species) tissue. Plant tissue was ground manually in extraction bags using a hand-held device before using as input for each method. For high-throughput automated purification, the Maxwell® RSC Instrument and KingFisher™ Flex Instrument were used. These platforms utilize DNA-binding paramagnetic particles to isolate and purify DNA. For low-throughput manual purification, a column-based method was used. Phytoplasma DNA purified using each method was successfully detected via qPCR using primers specific to the well-conserved 16S rDNA of phytoplasmas. The various purification methods and detection method described here are applicable to a broad spectrum of plant pathogen applications.

PO0211: Microbes and Pathogens

The "OUTPACE" Phytopathology Summer Institute: Taking Plant Molecular Biology to an Urban Garden

Karolina Mukhtar, University of Alabama at Birmingham, Birmingham, AL

OUTreach Plant PATHology Clinic & Education (OUTPACE) is a summer research program at the University of Alabama at Birmingham funded by the National Science Foundation CAREER award. OUTPACE is designed to provide undergraduate students from a largely African-American, urban setting with valuable research experience in plant molecular biology, knowledge about plant pathology and microbial laboratory techniques, in tandem with a service learning experience benefiting urban gardening families.

The OUTPACE 7-week-long summer curriculum consists of lectures, labs and fieldwork. Lectures offer the necessary background in theoretical plant pathology, while laboratories cover specific phytopathogen infection techniques using *Arabidopsis* and *Nicotiana benthamiana*. An integral part of OUTPACE is the "UAB Plant Clinic" initiative, where the participants work with the gardeners to identify and diagnose infected plant material in the UAB Community Gardens (service learning component). The final stage of OUTPACE is data analysis, interpretation of the results and preparation of lab reports, as well as the final reflection of the service learning experience and compiling the collections of newly identified pathogen strains/isolates. Graduate students from the PI's laboratory serve as peer mentors and gain teaching experience. The growers in the Community Gardens are engaged in citizen science through recording disease-related observations in their plots, providing descriptions and photographs, and communicating their findings to the PI throughout the year via an outreach URL.

We report on five years of the OUTPACE program. Since its inception in 2014, the program has trained 49 undergraduate students. Typical demographic make-up in a given year was 60-75% females and 30-40% ethnic minorities. The program alumni achieved a 100% retention and 100% 4-year college graduation rates. Rates of successful applications to biology- and biomedical-related professional and graduate programs were ~55% higher than in the general student population. Progress in attitudes and self-perception was measured using pre- and post-assessment surveys, and demonstrated highly positive outcomes.

The students report on their progress and learning activities online in a form of a scientific blog, available at <https://outpaceuab.wordpress.com/>.

In summary, the OUTPACE program is a successful example of a broader impact and outreach initiative that benefits undergraduate and graduate students, local citizens and community, and the PI's research and educational programs. These outcomes are in line with current regional and national initiatives to build healthy communities, promote civic engagement and sustainable living, and combat urban malnutrition and hunger.

PE0212: Microbes and Pathogens

Field-Based Real-Time Detection of Tick and Rodent-Borne Pathogens using Nanopore Sequencing

Laramie L. Lindsey¹, Nusrat Jahan¹, Kenwyn Shriner¹, Faisal A. A. Khan² and Peter A. Larsen¹,
(1)University of Minnesota, Saint Paul, MN, (2)University of Malaysia at Sarawak, Sarawak, Malaysia

The miniaturization of PCR and nanopore sequencing technologies allow for the outfitting of mobile genomic laboratories that can be deployed almost anywhere in the world. These technological advances have ushered in an exciting new era of field-based research, especially in the realm of molecular diagnostics and biosurveillance of emerging zoonotic pathogens. In the summer of 2019, we deployed a mobile sequencing lab to both the forests of Minnesota and the jungles of Borneo, Malaysia. Our goal was to determine the utility and flexibility of field-based nanopore sequencing for the real-time surveillance of both vector-borne and rodent-borne zoonoses that are of concern to the global One Health community. In Minnesota, we screened individual ticks collected from a state park for the presence of Lyme disease (causative agent *Borrelia burgdorferi*). In Borneo, Malaysia we conducted a week-long PCR and nanopore sequencing educational workshop at a wildlife preserve and, at the same time, we generated metagenomic sequence data from rodents collected near picnic areas across the preserve. Results of our field-based research efforts include the sequencing and phylogenetic placement of *Borrelia burgdorferi* strains in less than 12

hours and the discovery of *Listeria spp.*, *Vibrio spp.*, *Yersinia spp.*, *Salmonella spp.*, *Leptospira spp.*, and *Burkholderia spp.* at the rodent-human interface in Borneo. Collectively, our research helps to document the potential for nanopore sequencing-based mobile laboratories in the global fight against emerging zoonotic disease.

PO0213: Microbes and Pathogens

Rapid Identification of Foodborne Pathogens from the Rodent-Agricultural Interface using Nanopore Sequencing

Nusrat Jahan, University of Minnesota, Saint Paul, MN

The effective control of rodent populations on farms is a critical component of food-safety, as rodents are reservoirs and vectors for many foodborne pathogens (e.g., *Salmonella spp.*, *E. coli* O157, etc.). The functional role of rodents in the amplification and transmission of foodborne pathogens is likely underappreciated in the United States. Clear links have been identified between rodent pests and outbreaks of foodborne pathogens throughout Europe and Asia, however, comparatively little research has been devoted to studying this Rodent-Agricultural interface in the USA, particularly across the Midwest. Here, we address this existing knowledge gap by characterizing the metagenomic communities of rodent pests collected from Minnesota dairy farms. We leveraged the Oxford Nanopore MinION sequencer to provide a rapid real-time survey of the putative food-borne pathogens. Our preliminary data suggests the presence of pathogenic strains of *Salmonella spp.*, *Listeria monocytogenes*, *Campylobacter spp.*, *Clostridium spp.*, and *Escherichia coli* O157, along with many important mastitis pathogens. A critically important observation is that we discovered these pathogens within five species of rodents (*Mus musculus*, *Rattus norvegicus*, *Peromyscus leucopus*, *Microtus pennsylvanicus*, and *Peromyscus maniculatus*). Data generated from our study will likely result in the identification of new reservoirs for food-borne pathogens and species-specific traits. It is critical to use a 'One Health' approach in order to prevent and control the spread of zoonotic pathogens in our food supply. Furthermore, knowledge gained from our research efforts will directly inform and improve upon farm-level biosecurity efforts and public health interventions to reduce future outbreaks of foodborne disease.

PE0214: Microbes and Pathogens

Integrating the Genetic and Physical Maps of *Bremia lactucae* (lettuce downy mildew)

Lin Zhang, Kyle Fletcher, Juliana Gil, Cayla Tsuchida, Weiqi Tan, Melodie Najarro, Charlotte Acharya, Keri Cavanaugh, Aubrey Kenefick, Kelsey Wood and Richard Micheltmore, The Genome Center, University of California, Davis, CA

Lettuce is the most valuable vegetable crop in the United States with an annual production worth over \$2.4 billion. Lettuce downy mildew, caused by *Bremia lactucae*, is the major disease affecting lettuce worldwide. Rapid variation in this oomycete has hindered successful deployments of resistance genes and fungicides, while high heterozygosity and repeat level have obstructed genomic analysis. A high-quality, near-chromosome-scale consensus assembly from multiple read types was recently published along with genome signatures of heterokaryosis in this highly variable oomycete species. However, the correct chromosome number remains unknown. A high-density genetic map would validate the genome assembly and provide the foundation for analyses of phenotypic variation. Therefore, we have analyzed the progeny of several sexual crosses of *B. lactucae* using whole genome shotgun sequencing. An ultra-high density genetic map was constructed using a pseudo-testcross phase-aware strategy. The current linkage map for the reference isolate SF5, contains 16 linkage groups spanning 1672 cM, with around 20 crossovers genome-wide per individual. Markers from another five populations derived from Californian, European, historical, and modern parents, are being aligned to the reference map. Draft genome assemblies of the parental isolates, as well as ultra-long ONT reads for three isolates are facilitating haplotype assembly, genotyping, and downstream analysis. This study will be the foundation for combining physical and genetic maps and revealing candidate genes for important traits, in order to dissect genetic and molecular mechanisms of virulence in this economically important downy mildew.

PO0215: Microbes and Pathogens

Development of a Library of *R-Avr* Gene Pairs for Lettuce and Downy Mildew

Anne K.J. Giesbers, Kelsey J. Wood, Lily M. Karim, Sharon Yang, Kyle Fletcher and Richard W. Michelmore, The Genome Center, University of California, Davis, CA

The most effective strategy to prevent yield losses of lettuce (*Lactuca sativa*) due to downy mildew (caused by *Bremia lactucae*) is the cultivation of resistant varieties. Deploying effective resistance (*R*) genes at the right time is crucial, because *R*-genes are often rapidly overcome by the pathogen. Plant pathogenic oomycetes secrete effector (*Avr*) proteins into plant cells to promote infection. These effectors manipulate plant processes, either through their own biochemical activity, or by interacting with plant target proteins. Recognition of effectors by *R* proteins in the plant can result in disease resistance. A library of *Avr-R* gene pairs would be beneficial for the data-driven spatiotemporal deployment of effective resistance genes. Furthermore, it would enable the stacking of multiple *R* genes to increase the evolutionary hurdle for the pathogen to become virulent. Effector genes of *B. lactucae* were predicted from transcriptomic and genomic data and have been transiently expressed in lettuce germplasm. Effector recognition by the plant leads to a hypersensitive response (HR) at the site of infiltration. Segregating populations of lines that respond differently to specific effectors have been used to map the genetic loci responsible for effector recognition. Isolation of the causal resistance genes from these genetically characterized resistance loci is challenging because of the presence of clustered candidate *R*-genes. We have identified an *Avr-R* gene pair that induced HR upon transient co-expression in lettuce. We are currently optimizing and implementing this pipeline for rapid identification and cloning of additional interacting gene pairs.

PE0216: Microbes and Pathogens

Genomic Diversity Analysis of Endophytic Microbiome in Banana (*Musa* spp.)

Simrandeep Singh, Shiva Abdollahi Aghdam, Rachel Lahowetz and Amanda MV Brown, Texas Tech University, Lubbock, TX

Plant microbiomes play a critical role in plant health and productivity. While greatly expanding the metabolic repertoire of plants, microbiomes also provide novel nutritional and defense pathways. Yet, the exploration of endophytic microbiomes associated with banana, an economically important and staple food crop, is in the embryonic phase. Besides a few 16S rRNA studies, there is limited knowledge of banana-associated microorganisms. This study presents the shotgun metagenomic analysis of the banana microbiome from the root, corm, and shoot tissues derived from *M. sikkimensis*, *M. textilis*, William's hybrid, FHIA-25, Dwarf Cavendish and two genotypes of *M. balbisiana*. By analyzing the average of 126.5 million genomic reads per microbiome using the metaWRAP pipeline, differences were observed in assembled microbiomes that depend on tissue type and plant genotype. As an example, leaf blobplots describe *M. textilis* dominated by Burkholderiales, Rhizobiales, and Pseudomonadales, whereas Enterobacteriales, Rhizobiales, and Sphingomonadales dominate *M. sikkimensis*. Root microbiome depicted a greater diversity, yet a few associations appeared specific to genotype. Interestingly, all tissues and genotypes harbored an abundant, AT-rich organism whose identity and function as a potential core banana-associated microbe is under analysis. Banana is being cultivated in more than 130 countries, making it one of the world's largest monocultures. However, farming practices and the dominance of the triploid Cavendish cultivar have led to the spread of pandemic diseases, most notably, 'Panama disease race-4'. We created an anti-microbial gene database to screen our data for potential disease-protective biosynthetic clusters to identify a protective microbiome inoculum against this disease.

PO0217: Microbes and Pathogens

Bioinformatic Analysis of Banana Disease-Protective Microbiome

Rachel Lahowetz, Shiva Abdollahi Aghdam, Simrandeep Singh and Amanda MV Brown, Texas Tech University, Lubbock, TX

Bananas (*Musa* spp.) are a major food and commercial product in over 130 countries. Banana plants are currently threatened worldwide by a number of diseases including *Fusarium oxysporum* f.sp. *cubense* race 4 (FOC4) (Panama disease). Many plants, including banana, contain a complex and stable microbiome that potentially imparts disease protection. A working hypothesis is that cultivation and domestication have altered plant microbiomes, possibly reducing disease resistance. This study is designed to uncover the core microbiota, core metabolome, and core metagenome through a multi-omics approach to determine the potential disease-resistant microbiome. The microbiome was extracted from above-ground (leaves and stem) and below ground (root and corm) from seven

varieties of banana with a wide genotypic range and sequenced with Illumina. Our bioinformatics pipeline maps reads to reference banana genomes to confirm enrichment, assembles, bins, annotates genes, and blasts them to a database of anti-microbial lactones, terpenes, bacteriocins, phosphonates, and indoles (~400 genes), and general defense functions such as polyketide synthase and siderophores (~28,000 genes). So far, 2.05 billion reads were obtained from DNA. Read mapping suggested that enrichment was successful. From the DNA assemblies, N50s ranged from 495 to 7080 and max contig length ranged from 7790 to 903511 bp, suggesting many contigs recover full-length genes. These results will address whether less-cultivated species have a more diverse microbiome that confers greater disease resistance from which a future protective inoculant might be developed.

PE0218: Microbes and Pathogens

Wheat Virus Identification using Oxford Nanopore Technology

John Fellers¹, Christian Webb², Madison Fellers³, Jessica Shoup-Rupp² and Erick DeWolf², (1)USDA ARS, Manhattan, KS, (2)Kansas State University, Manhattan, KS, (3)USDA-ARS, Manhattan, KS

Viral diseases are a limiting factor to wheat production. Viruses are difficult to diagnose in the early stages of disease development and are often confused with nutrient deficiencies or other abiotic problems. Immunological methods are useful to identify viruses, but specific antibodies may not be available or require high virus titer for detection. In 2015 and 2017, wheat plants containing *Wheat streak mosaic virus* (WSMV) resistance gene, *Wsm2*, were found to have symptoms characteristic of WSMV. Serologically, WSMV was detected in all four samples. Additionally, *High Plains wheat mosaic virus* (HPWMOV) was also detected in one of the samples. *Barley yellow dwarf virus* (BYDV) was not detected and a detection kit was not readily available for *Triticum mosaic virus* (TriMV). Initially, cDNA cloning and Sanger sequencing were used to determine the presence of WSMV; however, the process was time-consuming and expensive. Subsequently, cDNA from infected wheat tissue was sequenced with single-strand Oxford Nanopore sequencing technology (ONT). ONT was able to confirm the presence of WSMV. Additionally, TriMV was found in all of the samples and BYDV in three of the samples. Deep coverage sequencing of full length, single-strand WSMV revealed variation compared to the WSMV Sidney-81 reference strain and may represent a new variant which overcome *Wsm2*. These results demonstrate that ONT can more accurately identify causal virus agents and has sufficient resolution to provide evidence of causal variants.

PO0219: Microbes and Pathogens

Diversity of a Worldwide Population Sample of Wheat Powdery Mildew

Alexandros Georgios Sotiropoulos¹, Marion C. Müller¹, Beat Keller² and Thomas Wicker², (1)Department of Plant and Microbial Biology, University of Zurich, Zurich, Switzerland, (2)University of Zurich, Zurich, Switzerland

Background:

The spread of newly evolving pathogen strains is a major challenge for agriculture. *Blumeria graminis forma specialis tritici* (*B.g. tritici*) is the causal agent of wheat powdery mildew, one of the most important wheat diseases worldwide. *B.g. tritici* has a large and complex genome of which approximately 85% is derived from transposable elements (TEs). Previous studies have

highlighted the importance of recombination of ancient haplotypes in the evolution of new pathogen strains. We wanted to study the worldwide diversity and population structure of this fungus in order to identify its origin, its migration patterns and potential genes that might be important for its survivability.

Methods:

The genomes of more than 200 *B.g. tritici* samples from around the world were sequenced with Illumina technology. Using mostly SNP and INDEL data, we performed phylogenetic, admixture and demographic analyses along with scans for signatures of selection.

Results:

Phylogenetically, the different populations are clustering according to their geographical origin with the ancestral recombination graphs showing a clearer picture of the recombination patterns. Admixture analyses suggest seven ancestral populations, most of them coming from Asia, with isolates from Israel and China having the highest nucleotide diversity. Interestingly, admixture data suggest that human

activity and trade played an important role in the spread of pathogen strains and hybridization of mildew strains from Australia, America and Japan. The scans for signatures of selection showed some candidate sites of selection for more than one natural population.

Conclusions:

Israel is supported to be the origin of wheat powdery mildew, with admixture showing frequent gene flow between populations. Additionally, our data highlight the importance of human activity and trade in the spread and evolution of mildew strains. Further investigation is being done for candidate genes that are under selection.

PE0220: Microbes and Pathogens

Genome Sequencing and Analysis of *Diaporthe aspalathi* Isolate MS-SSC91 Causing Stem Canker in Soybean

Shuxian Li¹, Mona Tavakolan², Qijian Song³, Akash Chandra² and Nadim Alkharouf², (1)USDA-ARS, Crop Genetics Research Unit, Stoneville, MS, (2)Department of Computer and Information Sciences, Towson University, Towson, MD, (3)USDA-ARS, Soybean Genomics and Improvement Lab, Beltsville, MD

Diaporthe aspalathi (Syn. *Diaporthe phaseolorum* var. *meridionalis*) is the causal agent of the southern stem canker (SSC) disease in soybean. It is an important fungal pathogen in the *Diaporthe-Phomopsis* complex. The pathogen can kill plants from the middle to the end of the soybean growing season and cause severe yield loss. The mechanisms of pathogenicity causing SSC by *D. aspalathi* are not fully understood. Information about the genome structure of *D. aspalathi* is also lacking. In this study, we successfully assembled the draft genome sequence of a *D. aspalathi* isolate, designated MS-SSC91, that was isolated from the stem of a field-grown soybean plant in Mississippi, USA in 2006. The genome was estimated to be approximately 55 Mb in size with an overall G + C content of 51 %. Gene prediction analysis using the AUGUSTUS software trained with the parameters of the fungal species *Fusarium graminearum* identified 14,962 genes. The average sequence length of genes was 1,729 bp; the largest sequence length of a gene was 23 kb. Approximately 46% (25.8 Mb) of the whole genome sequence was in genes. Of the 25.8 Mb sequences within genes, 22.5 Mb were coding sequences. Blast2Go was then used to annotate

and assign Gene Ontology (GO) classifications to those genes. Genes were classified into the GO categories of biological process, molecular function and cellular component. RepeatMasker is currently being used to find repeats in the genome. Identification of horizontal transfer genes is also underway. The MS-SSC91 genome sequences will provide information on the genetic basis of fungal infection of the soybean stem. It is valuable for studying soybean-fungal interactions and developing new control strategies for this pathogen.

PO0221: Microbes and Pathogens

The Genome of the Soybean Cyst Nematode (*Heterodera glycines*)

Rick Masonbrink, Thomas Maier, Arun S. Seetharam, Thomas J. Baum and Andrew J. Severin, Iowa State University, Ames, IA

The soybean cyst nematode (*Heterodera glycines*) is a sedentary plant parasite that causes over billions in yield losses annually. It has spread across the soybean producing world, emerging as the primary pathogen of soybeans. The problem is exacerbated by *H. glycines* populations overcoming the limited sources of natural resistance in soybean and by the lack of effective and safe alternative treatments. Although there are genetic determinants that render soybean plants resistant to certain nematode genotypes, resistant soybean cultivars are increasingly ineffective because their multi-year usage has selected for virulent *H. glycines* populations. *H. glycines* relies on the comprehensive re-engineering of an infection site into a syncytium, as well as the long-term suppression of host defense to ensure syncytial viability. At the forefront of these complex molecular interactions are effectors, the proteins secreted by *H. glycines* into host root tissues. The mechanisms that control effector acquisition, diversification, and selection need to be understood before effective control strategies can be developed. As a foundation to obtain this understanding we developed a nine scaffold 158Mb pseudomolecule assembly of the *H. glycines* genome using PacBio, Chicago, and Hi-C sequencing. An annotation of 25,180 genes was predicted using a Mikado pipeline that utilized Repeatmasker, Braker, Trinity, Spades, Portcullis, Class2, Maker, and Stringtie informed by published short and long read expression data. Here we present the preliminary results from our assembly and annotation of *H. glycines*.

PE0222: Microbes and Pathogens

Characterization of the Pathogenicity Determinant(s) of Pepper Mottle Virus in *Nicotiana benthamiana*

Miao Fang and Kook-Hyung Kim, Seoul National University, Seoul, South Korea

Pepper mottle virus (PepMoV), a member of the genus *Potyvirus* genus, infects mostly *Capsicum* species including pepper and bell pepper that are important vegetable and spice crops in Korea. The incidence of PepMoV diseases on different pepper cultivars becoming prevalent throughout Korea and thus posing a threat to pepper production. We have previously collected 13 PepMoV isolates from nine regions comprising five provinces in Korea. Symptoms caused by these 13 isolates differed from inoculated indicator host plants including *Nicotiana tabacum* cv. X-nc and *N. benthamiana*. To further identify the responsible symptom determinant(s) of PepMoV, two out of 13 isolates including 134 and 205136 were used in this study. PepMoV isolate 134 causes severe necrosis and yellowing while 205136 causes mottle and yellowing symptoms on *N. benthamiana* plants. Using full-length infectious cDNA clone of PepMoV 134, several counterpart substitution and site-directed mutagenesis mutants have been constructed and inoculated onto *N. benthamiana* plants. Among the constructed counterpart substitution mutants, we observed that the mutant 4, which substituting NIB and CP region with that of 205136, induced the different symptoms compared to that caused by the isolate 134. In addition, the mutants 2.1 and 3.1, which substituting HC-Pro region and half of CI:Nla region replaced by counterpart regions of 205136 showed dramatically faster systemic movement and stronger viral accumulation than 134 on local and upper uninoculated leaves. The amino acid substitution at position 2773 (V2773D) of the NIB protein resulted in different symptoms without noticeably affecting systemic movement of the virus. Through the use of chimeric and amino acid substitution mutants, we determined the viral factors required for specific symptom development as well as for systemic infection of the PepMoV in *N. benthamiana*.

PO0223: Microbes and Pathogens

Whole Genome Sequencing of Putative Somatic Hybrids Between Formae Speciales of *Puccinia graminis*

Michelle N. K. Demers¹, Priyanka Surana², Davinder Singh¹, Peter Dracatos³, Peter Dodds⁴ and Robert Park³, (1)The University of Sydney, Sydney, NSW, Australia, (2)The University of Sydney - Plant Breeding Institute, Sydney, NSW, Australia, (3)University of Sydney, Sydney, NSW, Australia, (4)CSIRO Agriculture, Canberra, Australia

Puccinia graminis, the fungus which causes the cereal stem rust disease, is one of the most devastating plant pathogens globally. Isolates of *P. graminis* have been found in Australia that have highly unusual pathogenicity profiles and cannot be categorized into any known forma specialis. These isolates have been referred to as 'scabrum' rusts, and are suspected of being the result of a somatic hybridisation event between the wheat stem rust pathogen *P. graminis* f. sp. *tritici* (Pgt) and the rye stem rust pathogen *P. graminis* f. sp. *secalis* (Pgs).

It is postulated that *P. graminis*, which are dikaryotic, can exchange haploid nuclei between isolates and even between formae speciales. This would allow these fungi, which are strictly clonally-propagating in Australia, to undergo a large genetic re-assortment within a single generation. This would have large implications for the use of resistance genes introgressed from different species into host plants and for hybrid crops such as triticale. Recent evidence suggests that somatic hybridisation may be a more significant driver of genetic change in *Puccinia graminis* than previously thought, and so this study aims to compare the similarity between the genomes of Pgs and Pgt and determine the origins of the 'scabrum' rust and whether it originated from a somatic hybridisation event. To test this hypothesis, the first genomes of Pgs and the 'scabrum' rust have been assembled using PacBio long-read sequencing and assembled de novo, and the degree of similarity of the genome and transcriptome of the 'scabrum' rust to putative parents, Pgt and Pgs, will be determined.

Understanding the mechanisms affecting genetic variability of highly-mutable pathogens such as *Puccinia graminis* is vital to the success of cereal breeding programs worldwide, and will influence the choice of resistance genes to incorporate into new resistant cereal varieties. This project is currently underway, and the results of these experiments will be presented at the conference.

PE0224: Microbes and Pathogens

Pan-Genome-Wide Analysis of *Pantoea ananatis* Strains Identified Presence and Absence Variants Linked to Pathogenicity of Onion

Gaurav Agarwal¹, Divya Choudhary¹, Shaun P Stice², Ronald D Gitaitis¹, Brian H Kvitko² and Bhabesh Dutta¹, (1)University of Georgia, Tifton, GA, (2)University of Georgia, Athens, GA

Pantoea ananatis is a member of a *Pantoea* spp. complex that causes center rot of onion, a disease that significantly affects onion yield and quality in the United States and elsewhere. Although many strains of *P. ananatis* are pathogenic to onion, others are not, and of the latter, some are pathogenic on other hosts (e.g. *Eucalyptus* spp., maize, pineapple, and rice). The bacterium lacks Type II, Type III, and Type IV virulence-associated secretion systems, thus it is intriguing as to what host-mediated pathogenicity factors are associated with *P. ananatis* interactions in onion. A previous comparative genomics study with *P. ananatis*, using 10 strains, correlated bacterial virulence with four mega-plasmid borne genetic loci, which were termed the "onion virulence regions (OVR A-D)". In addition, a set of chromosomal genes also were associated with onion pathogenicity. To identify other factors associated with onion pathogenicity, we conducted a pan-genome wide analysis with a larger number of strains (n=81) possessing varying levels of aggressiveness determined by the onion scale assay. Pan-genome analyses revealed a large core genome of 3,153 protein-coding sequences, which is shared among 81 strains and a flexible accessory genome of up to 5,065 genes. Our phylogenomic analysis based on the presence and absence variants (PAVs) could distinguish pathogenic vs. non-pathogenic *P. ananatis* strains. A pan-genome-wide association study conducted using PAVs and phenotyping data (aggressiveness on onion scale) identified potential candidate genes. The annotated functions of these genes included leucine biosynthesis, methyltransferases, carbon-nitrogen metabolism, dephosphorylation of phosphorylated signal transducing proteins. The pan-genomic differences not only aided in differentiating onion-pathogenic from non-pathogenic strains but also provided potential evidence of constantly evolving accessory genes in *P. ananatis*.

PO0225: Microbes and Pathogens

Host-Induced Gene Silencing (HIGS) for Reduction of Aflatoxin Contamination in Corn

Ali A. Ahmed, UC DAVIS, DAVIS, CA, Ioannis Stergiopoulos, University of California Davis, Davis, CA and Richard Micheltore, The Genome Center, University of California, Davis, CA

Aflatoxins are a major contaminant of several food crops including corn and it is a global food safety hazard. Host-Induced Gene Silencing (HIGS) has been successfully used to silence genes in several plant pathogens. In this study, we employ HIGS to reduce *Aspergillus* growth and aflatoxin biosynthesis by silencing two sets of genes in *Aspergillus flavus* and *A. parasiticus*. Accordingly, two RNAi constructs were made and transformed into corn. The first targets two essential genes (Esse12). The second construct targets three aflatoxin biosynthesis genes (Afla123). We obtained T1 seeds from 120 (71 and 49) primary transgenic plants which were regenerated from 35 (20 and 15) events for Esse12 and Afla123 constructs, respectively. The constructs contain Bar resistance marker. In order to identify transgenic plants, 50 seeds/plant/event were sown for all 35 events and 10 days old seedlings were sprayed with glufosinate herbicide. As a result, 34 events segregated 1:1 and one event 3:1 (transgenic hemizygous: non-transgenic) indicating single and two locus integration, respectively. Using PCR, 31 events (17 Esse12 and 14 Afla123) confirmed to contain the RNAi hairpin transgene, whereas 4 events contain the Bar gene but not the RNAi transgene. Herbicide-resistant hemizygous T1 plants were selfed and mature kernels were harvested. 50 T2 seeds/event were sown and seedlings were sprayed with herbicide. Except for one event, all segregated 3:1 (transgenic: non-transgenic) as expected. Transgene homozygous plants were identified for all 31 events using Quantitative PCR. Preliminary *Aspergillus* infection assays were conducted on semi-mature kernels and cobs. As a result, 2 plants belonging to the same Esse12 event showed a reduction in fungal growth compared to non-transgenic control plants. More *Aspergillus* infection assay is underway. Promising lines containing the two different RNAi targets will be crossed in order to pyramid individual RNAi effects and results in stronger reduction in *aspergillus* growth and aflatoxin biosynthesis.

PE0226: Microbes and Pathogens

Progress in the Control of Peanut Smut Disease

Renee S. Arias¹, Alicia N. Massa¹, Marina Bressano², Francisco de Blas³, Alejandro M. Rago⁴, Mario I. Buteler⁵, Marshall C. Lamb⁶, E. Cinthia Conforto⁷, Juan Soave⁸, Victor S. Sobolev¹ and Brian E. Scheffler⁹, (1)USDA-ARS National Peanut Research Laboratory, Dawson, GA, (2)Universidad de Cordoba, Cordoba, Argentina, (3)IMBIV-CONICET, Cordoba, Argentina, (4)IPAVE-CIAP-INTA, Cordoba, Argentina, (5)Criadero El Carmen, General Cabrera, Argentina, (6)USDA-ARS- National Peanut Research Laboratory, Dawson, GA, (7)INTA-Argentina, Cordoba, Argentina, (8)Criadero El Carmen, Cordoba, Argentina, (9)USDA-ARS, Stoneville, MS

Peanut smut disease can cause up to 50 % loss in Argentina where the disease is endemic. Though peanut smut has not been reported in the United States, the USDA-ARS-National Peanut Research Laboratory has worked in collaboration with several research institutions in Argentina to develop molecular tools to understand both, plant resistance and the genetics of the pathogen. Introgression of resistance was studied in 94 recombinant inbred lines (RILs) from wild diploid peanuts, and 45 lines resulting from crosses with resistant landraces. In both cases, molecular markers linked to disease resistance were identified. The pathogen, *Thecaphora frezii* has shown resistance to most fungicides used in the peanut crop. We sequenced the 123,773 bp mitochondrial genome and several nuclear genes of *T. frezii* (*succinate dehydrogenase*, *ergosterol biosynthesis*, *cytochrome p450* and *beta tubulin*) that are target of fungicides applied to the peanut crop. This not only provided the basic molecular tools to study the population genetics of *T. frezii*, but also detected mutations that confer resistance to strobilurin fungicides. The genetic diversity of *T. frezii* is still in progress; whereas a smut resistant variety has been released.

PO0227: Microbes and Pathogens

Transcriptome Analysis of the Peanut-*Aspergillus* Interaction

Alicia N. Massa, Renee S. Arias, Victor S. Sobolev, Valerie A. Orner, Travis E. Walk and Marshall C. Lamb, USDA-ARS National Peanut Research Laboratory, Dawson, GA

Soil fungi, *Aspergillus* spp., are opportunistic pathogens that invade peanut seeds causing accumulation of aflatoxins. Limited understanding of the peanut-*Aspergillus* interaction and aflatoxin accumulation has hindered the development of resistant cultivars or alternative methods of control. This study explores gene expression changes in peanut seeds at early stages of infection with *Aspergillus flavus* NRRL 3357. A simultaneous RNA sequencing approach was used to capture both seed and pathogen specific transcripts. The experiments used two aflatoxin-resistant genotypes of a wild diploid *Arachis* species and a susceptible cultivated peanut. For each experiment, gene expression analyses and identification of differentially expressed genes were based on paired-end sequence reads of two biological replicates. Genes involved in phytoalexin biosynthesis, including stilbenes, were among the differentially expressed genes in aflatoxin-resistant genotypes, but not in the susceptible peanut cultivar. Mapping of reads to the aflatoxin biosynthesis gene cluster indicated that peanut regulates transcription of mycotoxin synthesis genes in *A. flavus*. This research is part of a joint effort to identify the genetic mechanisms underlying aflatoxin resistance.

PE0228: Microbes and Pathogens

Gene Expansion in the Tylenchomorpha: Detecting Genes Involved in Plant Host Rewiring

Rick E. Masonbrink, Thomas Maier, Thomas J. Baum and Andrew J. Severin, Iowa State University, Ames, IA

Parasites survive and reproduce by harvesting nutrients from their hosts, thereby exerting a toll on fecundity and prompting host selection. Eventually the host develops resistance, and these same selective forces are mirrored in the parasite until resistance is overcome. These ongoing cycles of reciprocal adaptation between host and parasite have a genomic signature in the form of gene duplications, which are frequently concurrent with these evolutionary outcomes. We studied this phenomenon via 11 publicly available genomes in the Tylenchomorpha, a clade of economically damaging plant parasitic nematodes. Using gene orthology and phylogenetics we discern gene expansions from three phylogenetic nodes: the Tylenchomorpha, a basal node segregating migratory nematodes from sedentary root knot and root cyst nematodes; the Tylenchoidea root-knot and cyst nematodes that have massively different modes of infection with similar consequences; and the Heteroderidae, the divergence of cyst nematode genera in *Heterodera* and *Globodera*. We then further our understanding of secreted gene evolution in the development of a feeding site for *Heterodera glycines*, by leveraging published expression datasets characterizing pre-parasitic and parasitic stages with differing aspects of host compatibility. Altogether, using a particular emphasis on genes with conserved secretion signals in multiple species, we examined the expansion of gene families across these clades to discern conserved mechanisms of host cell manipulation and parasitism in *H. glycines* and provide a resource for developing new modes of nematode resistance.

PO0229: Microbes and Pathogens

Apple Scab Resistance of *Malus floribunda* 821 Can be Broken Down By New North American Isolates of *Venturia inaequalis*

David Papp, Cornell AgriTech, Geneva, NY

Apple scab, caused by *Venturia inaequalis* (Cke.) Wint., is a destructive fungal disease of major apple cultivars worldwide. Thus, development of scab resistant cultivars is one of the highest priorities of apple breeding programs. The principal source of resistance for breeding programs has been the scab resistance gene *Rvi6* that originated from the Japanese crabapple *Malus floribunda* (Sieb.) sel. 821. Isolates of *V. inaequalis* able to overcome *Rvi6* have been identified in Europe, but have not yet been reported on the American continents. We recently discovered scab infection on *M. floribunda* 821 trees in a research orchard at Geneva, New York, USA, where approximately 10% of the leaves bore profusely sporulating apple scab lesions, many of which had coalesced to cover entire leaves. Chlorosis and pinpoint pitting symptoms typical of failed infections by *V. inaequalis* on hosts bearing the *Rvi6* and *Rvi7* genes were also observed. We assessed genetic diversity and population genetic structure of six *V. inaequalis* isolates collected from *M. floribunda* 821, one isolate from ‘Nova Easygro’, one isolate from ‘Golden Delicious’ and two isolates from Europe using 16,321 genome-wide SNPs. Population genetic structure and PCA separated the isolates into distinct European and USA groups. The foregoing suggests that the new *Rvi6* virulent isolates emerged within USA populations, rather than being transported from Europe. The overcoming of resistance in *M. floribunda* 821 but not in descendant cultivars suggests that durable resistance to apple scab will require a more comprehensive understanding of *Rvi6* mediated resistance in diverse genetic backgrounds.

PE0230: Microbes and Pathogens

Citrusgreening.org: An Open Access Portal and System Biology Resources for the Tritrophic Disease Complex

Mirella Flores-Gonzalez¹, Prashant S Hosmani¹, Noe Fernandez-Pozo², Susan Brown³, Lukas A. Mueller¹ and Surya Saha¹, (1)Boyce Thompson Institute, Ithaca, NY, (2)University of Marburg, Marburg, Germany, (3)Kansas State University, KSU Bioinformatics Center, Manhattan, KS

We have created an open access web portal with pathosystem-wide resources and bioinformatics tools for the host citrus (*C. clementina* and *C. sinensis*), vector asian citrus psyllid (*D. citri*) and its endosymbionts, and multiple pathogens including *Ca. Liberibacter asiaticus*. This portal is a holistic framework to enable us to understand pathogenesis and disease transmission. The endeavor will connect and mine data sets generated by the community to study the citrus greening disease complex.

The portal contains a variety of tools for omics data. DiaphorinaCyc and CitrusCyc are metabolic pathway databases that can be used to visualize transcriptomics and proteomics results to highlight pathways with differentially regulated genes. Psyllid Expression Network (PEN) contains expression profiles of *D. citri* genes from multiple life stages, conditions and hosts. Citrus Expression Network (CEN) incorporates public expression data for citrus from NCBI SRA. The portal also includes raw and annotated data from electrical penetration graph (EPG) recordings of ACP feeding on citrus and close relatives. Apollo, a user-friendly manual curation tool allows the research community to continuously improve this knowledge base as more experimental research is published. All tools like JBrowse, Biocyc, BLAST, CEN and PEN connect to a central database containing gene models for citrus, *D. citri*, endosymbionts and multiple *Liberibacter* pathogens.

We welcome submission of relevant data sets to enable sharing and allow the community to analyze data in an integrated system. Bulk downloads are available for all genome and annotation datasets from the FTP site (<ftp://ftp.citrusgreening.org>). The portal can be accessed at <https://citrusgreening.org/>.

PO0231: Microbes and Pathogens

Studying the Bacterial Diversity of the Rumen at High Resolution

Ruth Dayana Hernandez Reyes, Universidad de los Andes-Agrosavia, Bogota, Colombia

Ruth Hernández, Universidad de los Andes-Agrosavia

Complex microbial communities have the challenge that no single method is capable of characterizing all the microbial diversity present in the sample. Traditional metabarcoding (16S rRNA gene amplicon) or shotgun sequencing give different overviews of the most abundant bacteria, and a deeper view requires a significantly higher sequencing effort, increasing costs. We aimed to develop a methodology that allows the recovery of a larger variety of taxonomic groups in the rumen of cattle beyond what the traditional methods were capable of. We used a sucrose density gradient, a simple and cost-effective method, that allowed to separate the community of the rumen bacteria in eight different fractions by size and density. We performed an experimental design with three cows of the Colombian Bon breed and a bull of the Holstein breed to test if our methodology works in the rumen bacteria in bovines regardless of their sex or breed. We took samples of ruminal fluid while the animals were fasting and one hour after the animals were fed on grass. We performed 16S rRNA gene Amplicon libraries for each fraction of the gradient and for each cow. The results showed that some taxonomic groups are enriched in certain fractions of the gradient in comparison with a traditional total sample of ruminal fluid, which didn't undergo separation by a sucrose gradient. The smallest fractions of the gradient showed enrichment in small size bacteria like some families of the Bacteroidetes phyla, the Tenericutes and Mollicutes phyla and bacterial taxonomic group SCRI. Prevotellaceae and Lachnospiraceae were enriched in the middle fractions of the gradient for all the animals, while, the families Mogibacteriaceae, Coriobacteriaceae, Veillonellaceae and some families of the Clostridiales order were enriched in the largest fractions of the gradient. PERMANOVA analysis showed with statistical significance that there were differences in the composition of the bacteria in the different fractions of the sucrose gradient. The distances established by the PERMANOVA test showed that the smallest fractions of the sucrose gradient (5, 10 and 20%) were more similar to each other than the rest of the fractions in the gradient and the total sample of ruminal

fluid. The enrichment of taxonomic groups in sucrose gradient fractions which were not abundant in a conventional DNA extraction, led us to perform shotgun sequencing in the enriched fractions, in order to assemble, potentially, full genomes from the enriched metagenomes, thus characterizing novel bacterial species that were underrepresented from the total fraction of the sample. In this way it is possible to deepen in the knowledge of the genome of these microorganisms and their role in the ruminal ecosystem.

Submitter's E-mail Address: rd.hernandez@uniandes.edu.co

Session Selection: Microbes and pathogens

First author: Ruth Hernández

Presenting Author

Ruth Hernández

Universidad de los Andes-Agrosavia

Email: rd.hernandez@uniandes.edu.co

PE0232: Microbes and Pathogens

Detection and Identification of Heterodera Species with the MatMaCorp Solas 8 Platform

Heather Piscatelli, MatMaCorp, Lincoln, NE

Heterodera is a genus of globally distributed plant-parasitic nematodes that feed on roots of many economically important crops. Infective juveniles penetrate roots and develop a feeding site where ultimately mature females transform into a cyst containing hundreds of eggs. Cysts may survive in the soil for many years without a host. Egg hatch is stimulated by a combination of environmental cues and host-specific exudates. This host-specificity allows for crop rotation to be used in nematode management. However, there are now four different cyst species that exist in the central and western plains of North America. Effective management requires timely and accurate species identification. We have developed a custom test to simultaneously detect and identify all four of these economically important cyst species. First, DNA is isolated from a single nematode or cyst using a field-friendly method which requires no pipettor or centrifuge. Next, the DNA template is added to a lyophilized C-SAND assay that includes sequence-specific fluorescent probes for each of the four species. Finally, the reactions are processed and immediately analyzed on a Solas 8 portable device. In about two hours, we are able to detect and identify *Heterodera avenae* (cereal cyst), *H. glycines* (soybean cyst), *H. medicaginis* (alfalfa cyst), and *H. schachtii* (sugarbeet cyst) from soil or root samples, and differentiate them from *H. trifolii* (clover cyst) or other closely-related nematodes. This technology gives the user a tool to make a rapid diagnostic decision for pest management or for monitoring the distribution and dispersal of these invasive pest species.

PO0233: Microbes and Pathogens

A Genome-Wide Association Study of the Sugar Beet Pathogen *Cercospora Beticola* Identifies Novel Fungicide Resistance Mutations

Rebecca E Spanner, North Dakota State University, Fargo, ND

Cercospora leaf spot (CLS) is caused by the fungus *Cercospora beticola* and is the most destructive foliar disease of sugar beet worldwide. The sterol demethylation-inhibiting (DMI) fungicides are one of the most important tools for managing CLS. DMI fungicides bind to and inhibit the cytochrome P450 enzyme CYP51 required for synthesizing ergosterol which provides integrity to the fungal cell membrane. Quantitative resistance to DMI fungicides has emerged in *C. beticola* populations due to their repeated and widespread use. In previous studies, isolates with higher EC50 values overexpressed *CbCYP51* compared to DMI-sensitive strains. However, no causal mutation has been found linked to this expression change. In order to identify mutations responsible for DMI resistance in *C. beticola*, a genome-wide association study was carried out. Illumina paired-end whole genome re-sequencing was performed for 194 unique *C. beticola* strains sampled from different fields in the Red River Valley sugar beet growing region in 2016 and 2017. Their sensitivity to the DMI tetraconazole was phenotyped as EC50 values calculated via agar plate growth. Genome-wide association identified a significant locus on Chromosome 8 in close proximity to *CbCYP51*. Two mutations at this locus are our top candidates underlying this resistance. CRISPR-Cas9 genome editing is being developed for *C. beticola* to elucidate which mutations are contributing to DMI resistance.

PE0234: Microbes and Pathogens

Metagenomic Sequencing of Seven Hypervariable Regions with Ion Genestudio S5 in Sea Beet

Piergiorgio Stevanato, DAFNAE, Università degli Studi di Padova, Legnaro (Padova), Italy

Metagenomics sequencing rely on the analysis of 16S ribosomal RNA sequences. Some of these regions are highly conserved and used for taxa identification, while the more variable regions serve to identify genus or species. The choice of 16S RNA region can significantly affect the accuracy in the estimate of taxonomic diversity. In this study, we present results coming from Ion GeneStudio S5 technology using 16S Ion Metagenomics Kit to amplify seven hypervariable region, V2, V3, V4, V6, V7, V8 and V9, of bacterial 16S rRNA in sea beet (*Beta vulgaris* spp. *maritima*). We analysed leaf samples of sea beets seedlings growing in 12 sites of Adriatic coast. The on-line pipeline Ion Reporter provides simple tools to manage, analyse and track samples. We highlighted the key features of Ion GeneStudio S5 sequencing technology together with Ion Reporter software for metagenomic 16S sequencing and data analysis. These information could be fundamental in guiding the choice towards a specific technology for data sequencing and analysis.

PO0235: Insects

Every Species Can be a Model: Reference-Quality PacBio Genomes from Single Insects

Sarah B Kingan¹, Christine Lambert¹, Liuqi Gu², Katharina von Wychetzk³, Primo Baybayan¹, Konrad Lohse⁴, James R Walters², Mara Lawnczak³ and Jonas Korlach¹, (1)Pacific Biosciences, Menlo Park, CA, (2)University of Kansas, Lawrence, KS, (3)Wellcome Sanger Institute, Hinxton, United Kingdom, (4)Institute of Evolutionary Biology, Edinburgh University, Edinburgh, United Kingdom

A high-quality reference genome is an essential resource for primary and applied research across the tree of life. Genome projects for small-bodied, non-model organisms such as insects face several unique challenges including limited DNA input quantities, high heterozygosity, and difficulty of culturing or inbreeding in the lab. Recent progress in PacBio library preparation protocols, sequencing throughput, and read accuracy address these challenges. We present several case studies including the Red Admiral (*Vanessa atalanta*), Monarch Butterfly (*Danaus plexippus*), and *Anopheles malaria* mosquitoes that highlight the benefits of sequencing single individuals for de novo genome assembly projects, and the ease at which these projects can be conducted by individual research labs. Sampled individuals may originate from lab colonies of interest to the research community or be sourced from the wild to better capture natural variation in a focal population. Where genomic DNA quantities are limited, the PacBio Low DNA Input Protocol requires ~100 ng of input DNA. Low DNA input samples with 500 Mb genome size or less can be multiplexed on a single SMRT Cell 8M on the Sequel II System. For samples with more abundant DNA quantity, size-selected libraries may be constructed to maximize sequencing yield. Both low DNA input and size-selected libraries can be used to generate HiFi reads, whose quality is Q20 or above (1% error or less) and lengths range from 10 – 25 kb. With HiFi reads, de novo assembly computation is greatly simplified relative to long read methods due to smaller sequence file sizes and more rapid analysis, resulting in highly accurate, contiguous, complete, and haplotype-resolved assemblies.

PE0236: Insects

Low-Cost Genomic Resources for Agricultural Pests

Erin D. Scully¹, Scott Geib², Nathan A. Palmer³, Sheina Sim², Scott Sattler³ and Gautam Sarath³, (1)USDA-ARS Center for Grain and Animal Health Research, Manhattan, KS, (2)USDA-ARS, Hilo, HI, (3)USDA-ARS Wheat, Sorghum, and Forage Research Unit, Lincoln, NE

Every year, insects cause major agricultural losses to crops in the United States and throughout the world. Although insecticides, crop rotations, natural enemies, mating disruption, and other approaches have eventually helped us overcome problems with insect pests, resistance to these management approaches is becoming increasingly common in many insect lineages and we have a stronger need than ever to understand how these resistance traits evolve in populations and seek alternative approaches for effectively controlling insect populations. Having access to high quality genome assemblies and gene annotations for these insect pests can help us develop new management approaches and potentially mitigate problems with resistance before it becomes widespread. Technologies for generating high quality genome assemblies from insects are rapidly advancing and are allowing us to release high

quality genome assemblies at unprecedented speed and scale. Over the past 3 years, USDA has adopted a number of approaches for producing high-quality genome assemblies from low DNA inputs, including 10X Chromium, PacBio, and Nanopore. Collectively, these approaches have led to the assembly of three major aphid pests of bioenergy grasses and >9 genome assemblies for stored product insects. These assemblies are currently being used to map mutations associated with cold tolerance and insecticide resistance as well as identify genetic factors that influence dietary niches.

PO0237: Insects

Genome-Wide Patterns of Genetic Differentiation of U.S. Commercial Honey Bee Stocks

Perot Saelao, Michael Simone-Finstrom, Arian Avalos, Lanie Bourgeois, Robert G. Danko, Lilia de Guzman, Frank Rinkevich@usda.gov and Philip Tokarz, USDA-ARS, Baton Rouge, LA

The genetics of U.S. honey bee stocks remain poorly characterized despite the importance of *Apis mellifera* as a crop pollinator. Several breeding programs have made significant improvements of favorable genetic traits. The variety of bees produced by artificial selection provides an exciting opportunity to explore the landscape of genetic diversity in commonly utilized stocks. Population genetic analyses found strong genetic similarity among seven stocks, while Pol-line, a stock with for mite resistance, showed significant differentiation likely due to strong selective breeding. Juxtaposing the underlying genetic variation of stocks selected for disease-resistance behavior, we identified genes and candidate regions potentially associated with resistance regulated by hygiene. This provides additional evidence for future studies towards understanding the genetic architecture of hygienic behavior. This study provides important insights into the distinct genetic characteristics and population diversity of honey bee stocks used in the United States. Composite signatures of selection helped highlight regions putatively under selection and potentially associated with disease resistance behavior. This study presents the initial effort towards effectively cataloging the standing variation within widely used honey bee stocks.

PE0238: Insects

HymenopteraMine: Genomic Data Mining Tools for Honey Bee and Other Hymenopteran Insects

Christine G Elsik, Amy T. Walsh, Deborah A. Triant, Md Shamimuzzaman and Justin J. Le Tourneau, University of Missouri, Columbia, MO

HymenopteraMine is a data mining resource for hymenopteran insects, accessible through the Hymenoptera Genome Database (HGD; <http://hymenopteragenome.org>). The goal of HymenopteraMine is to accelerate genomics analysis by enabling researchers without scripting skills to create and export customized annotation datasets merged with their own research data for use in downstream analyses. The newest release of HymenopteraMine (v1.4) contains Amel_HAv3.1, the new highly contiguous genome assembly of *Apis mellifera* (European honey bee). We have mapped the honey bee official gene set (OGSv3.2) to the new assembly, and computed cross-references between OGSv3.2 and RefSeq, facilitating meta-analysis between the old and new gene sets. For users completing studies with the older *A. mellifera* assembly (Amel_4.5), the previous release of HymenopteraMine (v1.3) is still accessible on the HGD home page.

In addition to *A. mellifera*, HymenopteraMine contains genomic data for 12 other bee species, 20 ant species, 10 wasp species and 3 sawfly species. The *Drosophila melanogaster* genome is also included so users can take advantage of well curated model organism data sets. With both simple and sophisticated search tools, users can create integrated data sets from NCBI, RefSeq, UniProt, InterPro, OrthoDB, KEGG, Reactome, Pubmed and Gene Ontology. Built-in template queries serve as starting points for data exploration, while the QueryBuilder tool supports construction of custom queries. With the List Tool, users can upload identifiers to build custom lists that can be analyzed with set operations; in the case of gene identifiers, the List Tool provides gene set enrichment for pathways and Gene Ontology. The Genomic Regions search tool executes queries based on uploaded lists of genome coordinates. HymenopteraMine facilitates cross-species data mining based on orthology and supports meta-analyses by tracking identifiers across gene sets. Query output can be exported in a variety of formats to be later used in further downstream analyses.

PO0239: Insects

Foam Bacterial Community Is Linked with the Gut of Nymphs of the Spittlebugs

Jose Mauricio Bento¹, Mateus Tonelli², Simone Raposo Cotta³, Alessandra Rigotto³, Armando C.F. Dias³ and Fernando Dini Andreote³, (1)University of Sao Paulo - ESALQ, Piracicaba, SP, Brazil, (2)University of Sao Paulo - ESALQ, Piracicaba, Brazil, (3)University of Sao Paulo, Piracicaba, Brazil

The development of insects is strongly influenced by their resident microorganisms. Symbionts play key roles in insect nutrition, reproduction and defense. Bacteria are important partners due to the wide diversity of their biochemical pathways that aid in the host development. We present evidence that the foam produced by nymphs of the spittlebug *Mahanarva fimbriolata* harbors a diversity of bacteria, including some that were previously reported as defensive symbionts of insects. Analysis of the microbiomes in the nymph gut and the soil close to the foam showed that the microorganisms in the foam were more closely related to those in the gut than in the soil, suggesting that the bacteria are actively introduced into the foam by the insect. Proteobacteria, Actinobacteria and Acidobacteria were the predominant groups found in the foam. Since members of Actinobacteria have been found to protect different species of insects by producing secondary metabolites with antibiotic properties, we speculate that the froth produced by *M. fimbriolata* may aid in defending the nymphs against entomopathogenic microorganisms.

PE0240: Insects

Allergen Profile and Metabolic Characteristics of *Tyrophagus putrescentiae* Revealed By Genome and Transcriptome Sequencing

Stephen Kwok-Wing Tsui¹, Qing Xiong¹, Zhi-Gang Liu² and Angel Tsz-Yau Wan¹, (1)School of Biomedical Sciences, The Chinese University of Hong Kong, Shatin, Hong Kong, (2)State Key Laboratory of Respiratory Disease for Allergy at Shenzhen University, School of Medicine, Shenzhen University, Shenzhen, China

Tyrophagus (T.) putrescentiae, also known commonly as a storage mite or mold mite, is an important mite species that infests a wide variety of stored foods, especially those with high protein and fat contents. *T. putrescentiae* is found to be associated with allergic disorders including asthma and allergic rhinitis similar to other domestic mites. In this study, both a high-quality reference genome and its whole-body transcription dataset of *T. putrescentiae* were constructed. Using a hybrid approach combining the Illumina and PacBio Sequel sequencing platforms, we were able to assemble a high-quality draft genome of 97.4 Mbp with only 176 scaffolds. Besides the 14 known allergens reported in public databases, 21 more possible allergen groups were identified by inferring homology from known allergens in other dust mites. Comparison of the metabolism and allergens in *T. putrescentiae*, *B. tropicalis*, *D. farinae*, and *D. pteronyssinus* was performed to uncover the differences in their habitats and allergenicity. Our results clearly demonstrated distinct allergen profiles in different mite species. Moreover, enzymes specifically found in *T. putrescentiae* could explain why this species is able to use fungus as the food source. In summary, here we report the first complete and accurately-annotated storage mite genome, which provides an important resource for the comparison in metabolic characteristics and allergenicity between *T. putrescentiae* and house dust mites.

PO0241: Insects

Applying Genome-Wide Markers to Identify the Origin and Demography of Boll Weevil Re-Infestations

Lindsey Perkin¹, Charles P.-C. Suh¹, Raul Ruiz-Arce², Tyler Jay Raszick³ and Greg A. Sword³, (1)USDA-ARS Southern Plains Agricultural Research Center, College Station, TX, (2)USDA APHIS Science & Technology, Edinburg, TX, (3)Texas A&M University, College Station, TX

The boll weevil, *Anthonomus grandis grandis*, has been eradicated from most of the cotton producing areas of the U.S., however, active populations in the Lower Rio Grande Valley (LRGV), TX and Mexico still pose a threat to eradicated areas. Post-eradication programs continue to operate to detect incipient weevil populations and re-infestations. Recent advancements in sequencing technology have resulted in many genomic and molecular tools that can now be applied to pest management. Herein, we demonstrate the use of some of these molecular tools to a recent boll weevil re-infestation in the Kingsville area of Texas. We used genome-wide SNP markers to determine the most likely geographical origin(s) of the weevils. We also used these markers to estimate the relatedness of the re-infestation population to find if weevils moved to the Kingsville via a natural mass dispersal from the LRGV or if

a few weevils were transported and then reproduced locally. Information to support the most likely scenario can be used by eradication programs to justify modifications to regulations in order to reduce anthropogenic movement of weevils.

PE0242: Insects

Genetic Variation Underlying Strain Specific Behavioral Response of *Tribolium castaneum* to Synthetic Lures

Alison R Gerken¹, James Campbell¹ and Erin D. Scully², (1)USDA ARS Center for Grain and Animal Health Research, Manhattan, KS, (2)USDA-ARS Center for Grain and Animal Health Research, Manhattan, KS

Stored product pest insects can infest and ruin millions of dollars in stored commodities each year. Synthetic lures using semiochemicals based on food and pheromone attractants have been developed to both monitor and manage pest populations. There can also be a variable response to these attractants within a given species and understanding the underlying causes of this variation is important for future formulations and use of these lures. *Tribolium castaneum*, the red flour beetle, has a rich set of genomic resources and a variety of independently isolated strains to examine variation in behavior and determine the genetic architecture of behavioral responses. Here we examined the attraction behavior of six strains of *T. castaneum* to both kairomone and pheromone lures in a wind tunnel. We found that attraction behavior among these six strains varied significantly, with generalized attraction to pheromone lures in 5 of 6 strains and extreme variation among individuals in one of the strains. Attraction to kairomone lures was more variable within and among the strains than attraction to pheromone lures. We then sequenced these six strains using Pool-Seq to find candidate regions or genes that may be involved in behavioral response to these attractants. These candidate genes have the potential to be functionally validated using RNAi and may provide targets for understanding behavioral genetic variation among stored product pests of a variety of species.

PO0243: Insects

The Comparative Genomics Analysis of House Dust Mites

Qing Xiong¹, Angel Tsz-Yau Wan¹, Malainual Nat² and Stephen Kwok-Wing Tsui¹, (1)School of Biomedical Sciences, The Chinese University of Hong Kong, Shatin, Hong Kong, (2)Department of Parasitology, Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok, Thailand

Background: House dust mites are the main cause of allergens causing human allergic diseases, especially respiratory diseases like asthma and allergic rhinitis. The most common species of house dust mites include *Dermatophagoides farinae*, *Dermatophagoides pteronyssinus* and *Blomia (B.) tropicalis*. Although house dust mites severely threaten public health, the associated biological background has not been well researched. The highly developed sequencing technology and comparative genomics techniques may therefore advance our knowledge to house dust mites.

Materials and Methods: A high-quality genome of *B. tropicalis* was assembled and well annotated. The genomes of other 20 species of ticks and mites and the mitochondrial genomes of other 39 species were downloaded from NCBI databases. MEGA7 was used for phylogenetic analysis, Orthofinder2 was utilized for gene family analysis and DAVID was employed for gene enrichment analysis.

Results: Overlapped BUSCO genes of 21 species of ticks and mites were aligned and constructed into a phylogenetic tree. As a species of house dust mites, *B. tropicalis* has a significant evolutionary distance from other house dust mites. In the comparative genomics analysis, proteolysis is the mostly enriched biological process term of *B. tropicalis*-specific orthogroups, which is resulted from the significant expansion of the chymotrypsin/trypsin protein family in the *B. tropicalis* genome through gene duplication. Also, a phylogenetic tree was constructed based on whole mitochondrial genomes of 40 species of ticks and mites.

Conclusions: The comparative genomics analysis paved a way to the comprehensive study to house dust mite and identified the gene family expansion of chymotrypsin/trypsin in the *B. tropicalis* genome. This finding is also consistent with the preferred high-protein diet of *B. tropicalis*.

PE0244: Insects

De Novo Assembly of a Chromosome-Level Reference Genome of *Periplaneta americana* Using Nanopore and Hi-C Sequencing

Lingyi Wang¹, Qing Xiong¹, Mai Shi¹, Angel Tsz-Yau Wan¹, Stephen Kwok-Wing Tsui¹ and Anchalee Tungtrongchitr², (1)School of Biomedical Sciences, The Chinese University of Hong Kong, Shatin, Hong Kong, (2)Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok, Thailand

Periplaneta americana, also known as the American cockroach, is one of the most common types of cockroaches in urban areas. With the ability to induce and spread infectious diseases in humans, it is long been recognized as an important cosmopolitan pest. In this project, we described a strategy of assembling chromosome-scale genome of American cockroach based on long-read sequencing and Hi-C technology. We also evaluated several popular assemblers with different subsets of raw reads in order to develop an optimal pipeline for assembly of large and complex insect genomes.

Nanopore raw reads were generated using GridION and PromethION 24. High-quality genome of American cockroach was generated using assembler Flye, Wtdbg2 or Ra followed by Illumina short reads polishing using Pilon. Then, Hi-C was performed, and HiRiseTM was used for correcting of misjoins and scaffolding.

The previously reported genome of American cockroach was assembled using short reads generated by Illumina sequencing technologies. Because of the high level of repetitiveness and diploid nature of the species, it was difficult to create a highly contiguous genome with short reads only. Here, taking the advantages of real-time long reads sequencing from Oxford Nanopore Technologies (ONT), we were able to obtain ultra-long reads with an average length of 10 kilobases which made the *de novo* assembly of 3.5 gigabases genome feasible. Polishing of the draft genome with nanopore raw reads and Illumina short reads increased BUSCO completeness from 96% to 98.3%. Hi-C further improved the genome assembly by correcting misjoins and improving order, orientation and contiguity of the genome up to chromosome-level. The first high-quality chromosome-level genome of American cockroach constructed and the assembly strategy used may serve as valuable resources for researchers who studying complex insect genomes.

PO0245: Insects

Identification of *Halyomorpha halys* Haplotype's Spread in Georgia By Mitochondrial DNA Sequencing

Mari Gogniashvili, Institute of Molecular Genetics, Agricultural University of Georgia, Tbilisi, Georgia and Nana Kunelauri, Tamar Shanava, Natia Tepnadze, Tengiz Beridze

The *Halyomorpha halys* is an invasive insect pest that attacks crop species and causes substantial economic damage. This insect family is native to China, Japan, Korea, and Taiwan where it has its biological enemy and therefore, doesn't spread uncontrollably. In Europe, it was first detected in Switzerland in 2004. It is now present in thirteen countries, and seems to be spreading throughout the continent. In 2015 it was spread in west Georgia, where it significantly damaged nuts, corn, and citrus seeds.

In the United States, Europe and Asia (10 Countries) 45 haplotypes were detected by mitochondrial genome (a 615 bp section of the *COI* gene) analysis. *H. halys* populations spread in Europe and North America from China with the exception of Greece, where *H. halys* was spread from Korea as well.

712 bp DNA mitochondrial cytochrome c oxidase I subunit gene fragment of *Halyomorpha halys* 17 samples spread in different regions of west Georgia were identified by sequencing. All Georgian *Halyomorpha halys* samples belong to H1 haplotype, which is native in Chinese population. It is possible that it spread either from Italy and Hungary, where *H. halys* is spread as a dominant haplotype or from Romania, where only this haplotype is widespread.

Complete nucleotide sequence of mitochondrial DNA of H1 haplotype of *H. halys* species was determined. Mitochondrial DNA sequencing was performed on an Illumina MiSeq platform. Mitochondrial DNA molecules were assembled using the SOAPdenovo computer program.

PE0246: Insects

Genomic Analysis of the H-Fibroin Gene in Trichoptera Silk

Adam M Taylor, Brigham Young University, Pleasant Grove, UT

Caddisfly (Insecta: Trichoptera) silk is unique from other insect's silk in that it retains its adhesive capabilities, strength and viscoelasticity when submerged in water. To understand how caddisfly silk is capable of possessing these characteristics, it is essential to understand the genetic foundation behind the silk proteins. Caddisfly silk is complex and made up of different structures generated by processes that are unique to caddisfly silk. H-Fibroin and L-Fibroin have been identified as two of the major protein components within caddisfly silk (Hatano & Nagashima, 2015). The H-Fibroin experiences unique post translational phosphorylation and has been found to contain high quantities of divalent metal ions. These unique structures of the silk contribute to the capabilities the silk possesses. An understanding of the primary structure of the protein is essential in understanding the complexity of the silk. In this study, we used next-gen sequencing technology to assemble the complex H-Fibroin gene in order to look at the underlying genetic structure of the protein. In doing so, we identified its unique repetitive sequence which contributes to Caddisfly silk's adhesive capabilities, strength and viscoelasticity when submerged in water.

PO0247: Insects

A High-Quality Reference Genome for *Diaphorina citri* using Single-Molecule Sequencing and Hi-C Proximity Ligation with Manually Curated Genes in Developmental, Structural and Immune Related Pathways

Prashant S Hosmani¹, Mirella Flores-Gonzalez¹, Teresa Shippy², Sherry Miller², Wayne Hunter³, Tom D'elia⁴, Susan J. Brown⁵, Lukas A. Mueller¹, **Surya Saha**¹ and International Psyllid Annotation Consortium, (1)Boyce Thompson Institute, Ithaca, NY, (2)Kansas State University, Manhattan, KS, (3)USDA Agricultural Research Service, Fort Pierce, FL, (4)Indian River State College, Ft. Pierce, FL, (5)Division of Biology - Kansas State University, Manhattan, KS

The vector, *Diaphorina citri* (Asian citrus psyllid) is the primary target of approaches to stop the spread of the pathogen *Ca. Liberibacter asiaticus* that causes Huanglongbing or citrus greening disease. Our focus is on creating better genomic resources to enable rapid functional discovery to target disease progression. The previous psyllid genome (Diaci v1.1) in NCBI was missing 25% of the single copy markers conserved in other Hemipterans. Manual genome annotations were able to identify a significant number of genome anomalies, misassemblies and missing genes. We present an improved *de novo* assembly with PacBio long reads followed by Dovetail Chicago and Hi-C based scaffolding. The current assembly (Diaci v3) has 13 chromosomal length scaffolds with a genome size of 475 Mb. Full-length cDNA transcripts were sequenced with PacBio IsoSeq technology from diseased and healthy tissue at multiple life stages. IsoSeq along with diverse Illumina RNA-Seq expression data were used to predict ~20K protein-coding genes in psyllid using MAKER annotation pipeline. We also generated genome independent transcriptome with a comprehensive catalog of all genes in the psyllid.

Gene-targeting technologies like RNAi, antisense oligos and CRISPR require accurate annotation of genes. Lack of closely related well characterized model organisms coupled with the diversity of insect genomes impacts the quality of predicted gene models. A high-quality manually curated gene set for developmental, structural and immune pathways is presented. Gene sets involved in chitin metabolism, cuticle, Hox/Hox cofactors, segmentation, chromatin remodeling complexes, circadian rhythm, carbohydrate metabolism, iron metabolism and endocytosis were manually curated using Apollo annotation editor. Immune related pathways with manually annotated genes include pathogen recognition molecules, signaling cascades associated with pathogenesis (Toll, IMD and JAK-STAT pathways) and response to pathogens. All resources will be available on <https://citrusgreening.org/> a portal for all omics resources for the citrus greening disease research community.

PE0248: Insects

Genomic Analysis between Trichoptera and Lepidoptera Show Evolutionary Innovations Allowing Trichoptera to Adapt to an Aquatic Environment

Lindsey Olsen, BYU, San Diego, CA

Trichoptera (caddisflies) have evolved to become the most diverse, exclusively aquatic insects, yet many of the genomic changes that contribute to the success of this order of insect are still unknown. Trichoptera and Lepidoptera (moths and butterflies) are reciprocally monophyletic meaning that they both share their most recent common ancestor. Despite being closely related, Trichoptera have evolved into the most diverse, exclusively aquatic insects, whereas, Lepidoptera have evolved to become a diverse, almost exclusively terrestrial insect (Holzenthall et al. 2007). Trichoptera and Lepidoptera are the subjects of scientific inquiry because they are both capable of spinning silk. Trichoptera produce silk as larvae and use it to make cases or fixed retreats. Trichoptera silk is of particular interest because its properties allow for it to be an underwater adhesive. While other research has focused primarily on the evolution of Trichoptera silk, little research has been done to identify the evolutionary innovations that allowed Trichoptera to adapt and diversify in an aquatic environment. Our research focuses on identifying the genomic basis of their evolutionary innovations. We report the genome annotation of four newly sequenced Trichoptera species *Hesperophylax magnus*, *Parapsyche elsis*, *Philanisus plebeius*, and *Rhyacophila brunnea*. These annotations reveal levels of homozygosity, conserved elements, and gene duplications. We then conducted a genome-wide search for gene family expansions and retractions using CAFE, in order to identify genomic regions that could contribute to Trichoptera's unique qualities and evolutionary history.

Holzenthall R. W., R. J. Blahnik, A. L. Prather, and K. M. Kjer, 2007 Order Trichoptera Kirby, 1813 (Insecta), Caddisflies*. Zootaxa 1668: 639–698. <https://doi.org/10.11646/zootaxa.1668.1.29>

PO0249: Insects

Molecular Characterization of Mosquitoes in North Central Nigeria Using Internal Transcribed Spacer 2 (ITS2), Mitochondria 16s rRNA and Intergenic Spacer Region (IGS)

Rahmat D. Shaibu, University of Ilorin, Ilorin, Kwara State., Nigeria, Oluyinka A. Iyiola, University of Ilorin, Nigeria and **Temitope O. Fadipe**, Federal Institute of Industrial Research, Oshodi, Lagos, Lagos, Nigeria

Mosquitoes are vectors of different diseases like malaria, filariasis, yellow fever etc. They are usually found around human habitations and this allows for easy transmission of the diseases from these vectors to human. Different species have been identified morphologically but this method is not very informative when identifying sibling species as they look very identical so molecular methods have been employed to characterize the different mosquito species and sibling species. This study was aimed at identifying and characterizing different mosquito species in five states in North-Central Nigeria using internal transcribed spacer 2 (ITS2) region, mitochondrial 16S rRNA and intergenic spacer (IGS) region. Over 3,000 mosquito species belonging to the genera *Anopheles*, *Culex* and *Aedes* were collected. Larval samples of these mosquitoes were collected from different sites in FCT, Kwara, Kogi, Benue and Niger and they were allowed to emerge to adults after which they were collected into containers using aspirators and put in the freezer and allowed to die. They were carefully placed in tubes with silica gel in them for preservation. DNA was extracted from the whole mosquito and the ITS2, mt16S rRNA and IGS regions were amplified. Analysis of sequences from the ITS2 region was able to distinguish two mosquito subfamilies; Anophelinae and Culicinae as well as differentiate between and amongst *Culex* and *Aedes* species. However, it was unable to effectively distinguish between the two different species of *Anopheles* sequenced. Mitochondrial 16S rRNA was also able to distinguish the two mosquito subfamilies. It efficiently identified and differentiated the different *Culex*, *Aedes* and *Anopheles* species sequenced for the study. IGS was only able to identify *Anopheles* species effectively. This study characterized the different mosquito species analysed in North-Central Nigeria and gave information about their evolutionary divergence.

PE0250: Insects

Population Structure and Impact on Resistance to Transgenic Crops in Fall Armyworm (*Spodoptera frugiperda*) from Diverse Locations

Katrina Schlum¹, Juan Luis Jurat-Fuentes², Kurt Lamour², Caroline Placidi de Bortoli² and Scott J. Emrich³, (1)University of Tennessee, Department of Genome, Science & Technology, Knoxville, TN, (2)University of Tennessee, Department of Entomology and Plant Pathology, Knoxville, TN, (3)University of Tennessee, Department of Electrical Engineering and Computer Science, Knoxville, TN

The fall armyworm (*Spodoptera frugiperda*, J.E. Smith) is a highly polyphagous agricultural pest with long-distance migratory behavior threatening food security worldwide. This pest has a host range of >80 plant species but preferentially feeds on corn and has developed resistance to multiple pesticides. Specifically, field populations of *S. frugiperda* in North and South America have evolved practical resistance to transgenic corn producing the Cry1Fa insecticidal protein from the bacterium *Bacillus thuringiensis* (*Bt*). The mechanism and allele linked to resistance against Cry1Fa in Puerto Rico was identified as a 2-bp insertion in an ATP binding cassette subfamily C2 (*ABCC2*) gene in *S. frugiperda* (*SfABCC2*). The goal of this project was to survey fall armyworm genomes from the USA, Brazil, Argentina, Kenya and Puerto Rico to determine population structure and how it may affect resistance evolution. Based on whole genome sequencing of 51 *S. frugiperda* samples, our analysis suggests there is no clear population structure based on BUSCO gene or k-mer based trees and principal component analysis. In contrast, two subpopulations were found based on a mitochondrial gene-based tree, supporting the continued use of mitochondrial gene markers (*COI* and *Tpi*) for surveying *S. frugiperda* populations. Our research is the largest diverse collection of *S. frugiperda* whole genome sequences characterized to date and provides a foundational resource for surveying *S. frugiperda* populations and resistance candidate alleles.

PO0251: Natural Populations

Towards Adding the Epigenome to Landscape-Level Genome-Phenotype Studies in Trees: Modeling and Empirical Approaches

Amy Whipple, NAU, Flagstaff, AZ, Ehren Moler, Northern Arizona University, Flagstaff, AZ, Erin Landguth, University of Montana, Missoula, MT and Samuel A. Cushman, USDA Forest Service, Rocky Mountain Research Station, Flagstaff, AZ

The potential of epigenomic modifications to influence phenotypes and be passed from parent to offspring is being shown in evermore plant and animal species. While, research in model systems in demonstrating the mechanisms by which epigenetic variation can persist, or not. How pervasive these mechanisms are, and what effects they have over time and space, remains to be seen. We have initial bsRADseq data of methylation in two individual trees in our southwestern white pine research system which has landscape genomic, environment-genome association, common garden, and genome-wide association studies underway. The methylation data is intended to add to our interpretation of the influence of temperature manipulations at different life stages on traits and whether the traits may be influenced by the epigenetic response to temperature. We also seek to gain more insight into the potential importance of epigenomic modifications via simulation models. CDMETAPOP is an existing, individual-based landscape-level simulation model. It models landscape genomic processes and includes neutral loci, as well as loci under selection according to spatial variation in environmental gradients. CDMETAPOP also incorporates varying landscape resistance to gene flow. Our first aim is to modify CDMETAPOP to accommodate the ways in which model systems have shown that epigenetic modifications may be modified, inherited and revert. Then simulations will be used to compare change through time in epigenetic and genetic contexts. Next, simulations of both genetic and epigenetic process will be modelled using parameters from a landscape genomic and common garden study of southwestern white pine, *Pinus strobiformis*. An eventual goal is to understand how these two evolutionary processes might interact to determine the phenotypic landscape.

PE0252: Natural Populations

Combining Citizen Science and Population Genomics to Reveal Global Invasion Routes: A Case Study with the Agricultural Pest *Pieris rapae*

Sean Ryan, Exponent, Inc., Menlo Park, CA, Eric Lombaert, Inra PACA, Sophia Antipolis, France, Anne Espeset, University of Nevada, Reno, Reno, NV, Roger Vila, Institut de Biologia Evolutiva (CSIC-UPF),

Barcelona, Spain, Gerard Talavera, Institute of Evolutionary Biology (CSIC-UPF), Barcelona, Spain, Vlad Dinca, University of Oulu, Oulu, Finland, Meredith Doellman, University of Notre Dame, South Bend, IN, Mark Renshaw, Oceanic Institute of Hawai'i Pacific University, Waimanalo, HI, Matthew Eng, Unified, New York, NY, Emily Hornett, University of Liverpool, Liverpool, United Kingdom, Yiyuan Li, University of Texas at Austin, Austin, TX, Michael Pfrender, University of Notre Dame, Notre Dame, IN and David DeWayne Shoemaker, University of Tennessee, Knoxville, TN

The small cabbage white butterfly, *Pieris rapae*, is a major agricultural pest of cruciferous crops and has been introduced to every continent except South America and Antarctica as a result of human activities. In an effort to reconstruct the near-global invasion history of *P. rapae* we developed a citizen science project, "Pieris Project," and successfully amassed thousands of specimens from 32 countries worldwide. We then generated and analyzed genomic (ddRAD) and mitochondrial DNA sequence data for these samples to reconstruct and compare different global invasion history scenarios. Our results bolster historical accounts of the global spread and timing of *P. rapae* introductions. We provide the first molecular evidence supporting the hypothesis that the ongoing divergence of the European and Asian subspecies of *P. rapae* (~1,200 yrBP) coincides with the diversification of brassicaceous crops and the development of human trade routes such as the Silk Routes (Silk Road). The further spread of *P. rapae* over the last ~160 years was further facilitated by human movement and trade, resulting in an almost linear series of at least four founding events, with each introduced population going through a severe bottleneck and serving as the source for the next introduction. Management efforts of this agricultural pest may need to consider the current existence of multiple genetically distinct populations. Finally, the international success of our citizen science project—Pieris Project—demonstrates the power of the public to aid scientists in collections-based research addressing important questions in invasion biology, and ecology and evolutionary biology more broadly.

PO0253: Natural Populations

Drift-Selection Interactions Drive the Diversification of *Pitcairnia lanuginosa* (Bromeliaceae) Across its Wide-Range and Structured Neotropical Distribution

Clarisse Palma Silva¹, Myriam Heuertz², Cleber Juliano Chaves³, Vanessa Araújo Graciano⁴, Luis Alberto Pillaca Huacre⁵ and Bárbara Simões Santos Leal³, (1)State University of Campinas, Campinas, Brazil, (2)INIA Forest Research Centre, Madrid, Spain, (3)Unicamp, Campinas, Brazil, (4)UNESP, Rio Claro, Brazil, (5)Museo de Historia Natural de la Universidad Nacional Mayor de San Marcos, Lima, Peru

Both genetic drift and divergent selection are predicted to be important drivers of species diversification within patchy habitats, but the extent to which these forces act on natural populations is strongly affected by species' ecological features. In this study, we infer the evolutionary history and genomic structure of *Pitcairnia lanuginosa*, a widespread species with a patchy distribution across Central Tropical South America. We sampled populations in the Brazilian Cerrado and Central Andean Yungas, and genotyped SNP markers defined through double-digest restriction-site associated DNA sequencing. In addition, we measured physiological traits and compared patterns of phenotypic (P_{ST}) and genetic (F_{ST}) divergence (P_{ST} - F_{ST} comparisons) in a subset of populations from the Cerrado. Our results from molecular analyses showed an extremely low genetic diversity, small effective population sizes and a remarkable differentiation among populations, supporting a major role of genetic drift for population divergence. In agreement with this observation, F_{ST} outlier tests and most P_{ST} - F_{ST} comparisons suggested a limited effect of selection as a force that would be driving the local adaptation of *P. lanuginosa* populations to mesic habitats across a wide distribution. However, P_{ST} - F_{ST} comparisons suggested divergent selection on few physiological traits linked to drought tolerance and Bayesian generalized linear mixed models revealed that genetic variation on outlier loci may be explained by isolation by environment in addition to isolation by distance among the Cerrado populations. Our study has important implications to improve our knowledge on the joint roles of genetic drift and divergent selection in generating divergence and diversity in species with a naturally patchy distribution.

PE0254: Natural Populations

High Throughput Genomic Resource Creation for Expediting Conservation Efforts in Medicinally Important *Trillium govanianum*

Praveen Dhyani, Balraj Sharma, Pradeep Singh, Mamta Masand, Sanjay Kumar and Ram K. Sharma, CSIR - Institute of Himalayan Bioresource Technology, Palampur, India

Trillium govanianum's ethnobotanical perspective includes the treatment of cancer, cardiovascular disease, neurasthenia, giddiness, arthritis, dysentery, and inflammation. The species is rich in steroidal saponins, a precursor for the synthesis of many steroidal drugs. Owing to its value for steroidal saponins there has been an illicit trade in the Indian Himalayan Region. However, in spite of its high medicative importance and threat to survival, conservation efforts in *T. govanianum* have been lacking, primarily due to non-availability of genome-wide genomic and molecular marker resources. Using next-generation tissue-specific transcriptome sequencing, comprehensive marker resource database comprising more than 7000 microsatellite markers were generated with an abundance of makers containing tri-repeats, primarily in the CDS region. Functional annotations with multifarious databases identified significant microsatellite containing transcripts putatively involved in various transcription factors and secondary metabolic pathways. Further, to evaluate the polymorphic potential of more than two hundred SSR primer pairs in random genotypes identified 102 polymorphic markers. This identified potential marker resource can be exploited for genetic diversity and genome mapping studies for identification of the elite populations for implementation of conservation strategies in *T. govanianum*.

PO0255: Natural Populations

Large Structural Variants Confound Measures of Introgression in Sunflowers

Gregory L. Owens, University of California, Berkeley, Berkeley, CA

Structural variants, including inversions, can prevent recombination and produce radically different evolutionary histories compared to the rest of the genome. Recent work has found numerous large (1-100 Mbp) structural variants segregating within three species of sunflower (genus *Helianthus*). Here we explore how these regions can produce divergent D-statistic values due to their evolutionary history. In particular, we focus on the subspecies, *H. annuus texanus*, which is thought to be of hybrid origin and is characterized by haplotype frequency changes at several structural variants.

PE0256: Natural Populations

The Roles of Host Tree Genomics and Environment in Driving Landscape-Scale Community Assembly

Helen M. Bothwell¹, Arthur R. Keith², Julia B. Hull², Samuel A. Cushman³, Catherine A. Gehring², Thomas G. Whitham² and Gerard J. Allan², (1)Australian National University, Canberra, ACT, Australia, (2)Northern Arizona University, Flagstaff, AZ, (3)USDA Forest Service, Rocky Mountain Research Station, Flagstaff, AZ

Rapid global climate change is causing widespread shifts in species distributions and community assemblages, yet few studies have investigated the relative contributions of host tree genomics and climate on associated community assembly at the landscape scale. We utilized a macrosystems approach to predict community consequences of climate change across southwestern riparian forests. Using genome-wide SNP data, we assessed genetic diversity and structure of Fremont cottonwood (*Populus fremontii*) across the southwestern US. For the same trees, we collected diversity and abundance data for associated leaf modifying arthropods and sequenced the leaf fungal endophyte communities. Four key findings emerged: 1) Fremont cottonwood is genetically differentiated into three major ecotypes, comprising the Sonoran Desert, Utah High Plateau, and California Central Valley. 2) Each ecotype occupies a significantly different, largely non-overlapping climate niche. 3) Both the leaf modifying arthropod and fungal endophyte communities were significantly differentiated across cottonwood ecotypes, with genetic distance among host trees explaining 13-24% of variation among associated communities, respectively. 4) After accounting for the effect of host tree genotype, climate emerged as the strongest predictor of community differentiation, with climate moisture deficit and precipitation seasonality contributing the greatest explanatory power. Our results suggest that host tree genetics is a critical component structuring associated communities at the local scale, however, climate is a stronger driver of community assembly at the landscape scale. To enhance resilience to climate change, we recommend planting for high genetic diversity locally, and stratifying restoration networks across the three ecoregions to maximize gamma diversity.

PO0257: Natural Populations

Natural Variation of Gene Regulatory Networks

Arthur Korte¹, Jan Freudenthal², William Lopez² and Ammarah Anwar², (1)Center for Computational and Theoretical Biology, University Wuerzburg, Wuerzburg, Germany, (2)CCTB, Würzburg, Germany

Understanding the causal relationship between genotype and phenotype is a major objective in biology. A default tool to illuminate these relationships are genome-wide association studies (GWAS). Here the goal is to identify genetic loci that associate with the trait of interest.

Genomic prediction (GP), on the other hand, aims to predict the phenotype from the genome. Both methods have been successfully used in many different species to elucidate trait architecture or prognose plant response. However, most studies concentrate on marginal marker effects and ignore epistatic and gene-environment interactions. These interactions are problematic to account for, but are likely to make major contributions to many phenotypes that are not regulated by independent genetic effects, but by more sophisticated gene-regulatory networks. A further complication arises from the fact that these networks vary in different natural accessions. Still, understanding the differences of gene regulatory networks and gene-gene interactions is crucial to conceive trait architecture and predict phenotypes.

I will present data on statistical approaches to tackle these challenges and present examples - using data from the Arabidopsis 1001 Genomes Project – of gene regulatory networks that have been realized differently in different natural accessions.

PE0258: Natural Populations

Local Adaptation, Genetic Divergence, and Experimental Selection in a Foundation Grass across the US Great Plains' Precipitation Gradient

Loretta C Johnson, Biology, Kansas State University, Manhattan, KS, Matthew Galliart, Kansas State University, Manhattan, KS, Nora Bello, Kansas State University, Department of Statistics, Manhattan, KS, Jesse Poland, Department of Plant Pathology, Kansas State University, Manhattan, KS, Paul St Amand, USDA-ARS-Small Grain Genotyping Lab, Manhattan, KS, Mary Knapp, Department of Agronomy, Manhattan, KS, Brian Maricle, Biology Department,, Hays, KS and Sara G Baer, Southern Illinois University, Carbondale, IL

This research seeks to understand factors that contribute to population genetic divergence, formation of ecotypes, and ultimately leading to new species. Habitats are often temporally and spatially variable, resulting in genetic divergence. *Andropogon gerardii*, dominant grass of the Great Plains, represents ~70% of biomass and has wide distribution across a sharp precipitation gradient. Ecotypes (dry, mesic, wet) were reciprocally planted as ecological communities in Colby, Hays, and Manhattan, KS, and Carbondale, IL. We tested for local adaptation over 6 years using single ecotype plots and plots with all three ecotypes mixed together. To analyze the genetic composition of the mixed ecotype plots, we used genotype information from plants of known ecotype, then trained a random forest model to assign unknown individuals to one of three ecotypes. Planting of ecotypes as a community and over multiple years is rarely done, but offers the most realistic test of local adaptation. PCA and STRUCTURE show strong genetic differentiation between ecotypes. Outlier analysis identified 64 markers under divergent selection, mainly related to rainfall. Candidate gene GA1, known to control internode length and height, was associated with strong height differences between ecotypes. These multi-year, community plantings indicate local adaptation to drought with the dry ecotype having higher cover in Hays, KS and the wet ecotype having greater cover in its home site of Carbondale, IL. Ultimately these results will provide recommendations to land managers on which climate-adapted source populations will be best suited for bioenergy and restoration plantings in future drier climates.

PO0259: Natural Populations

Impact of the Bastrop County Complex Fire on Soil Fungal Population across a Burn Severity Gradient

Claire Barnett, Antioch University, Seattle, WA, Greg Creacy, Texas Parks and Wildlife Dept., Bastrop, TX and **Charles Hauser**, St. Edward's University, Austin, TX

Globally, fire regimes are changing, with increases in wildfire frequency and severity expected for many North American forests. Fires can result in dramatic changes in plant and microbial communities, with long-lasting effects on ecosystem functions. Both natural and human caused fires are a common disturbance in the western and southwestern United States. We investigated the effects of the central Texas Bastrop County Complex Fire, that burned 34,330 acres including the Lost Pines ecosystem and Bastrop State Park, on soil fungal populations using high-throughput amplicon sequencing across a gradient of burn severity and soil types. Preliminary analysis suggests that fire intensity reduced alpha diversity (Shannon) as a function of burn severity, and PCA analysis identifies sub-clusters on the basis of burn severity. These preliminary results identify specific fire-responsive microbial taxa and suggest that accounting for burn severity improves our understanding of their response to fires, with potentially important implications for ecosystem function, and restoration efforts.

PE0260: Natural Populations

Loxodonta Localizer: A Software Tool for Inferring the Provenance of African Elephants and Their Ivory Using Mitochondrial DNA

Alfred L. Roca, University of Illinois at Urbana-Champaign, Urbana, IL

Illegal hunting is a major threat to the elephants of Africa, with more elephants killed by poachers than die from natural causes. DNA from tusks has been used to infer the source populations for confiscated ivory, relying on nuclear genetic markers. However, mitochondrial DNA (mtDNA) sequences can also provide information on the geographic origins of elephants due to female elephant philopatry. Here, we introduce the *Loxodonta* Localizer (www.loxodontalocalizer.org), an interactive software tool that uses a database of mtDNA sequences compiled from previously published studies to provide information on the potential provenance of confiscated ivory. A 316 bp

control region sequence, which can be readily generated from DNA extracted from ivory, is used as a query. The software generates a listing of haplotypes reported among 1,917 African elephants in 24 range countries, sorted in order of similarity to the query sequence. The African locations from which haplotype sequences have been previously reported are shown on a map. We demonstrate examples of haplotypes reported from only a single locality or country, examine the utility of the program in identifying elephants from countries with varying degrees of sampling, and analyze batches of confiscated ivory. The *Loxodonta* Localizer allows for the source of confiscated ivory to be assessed within days, using widely available molecular methods that do not depend on a particular platform or laboratory. The program enables identification of potential regions or localities from which elephants are being poached, with capacity for rapid identification of populations newly or consistently targeted by poachers. With 16 co-authors, shown on the first slide.

PO0261: Natural Populations

Mallard Introductions to New Zealand Result in Extensive Hybridization with Endemic Grey Ducks

Joshua I. Brown¹, Jonathon Mohl¹, Jennifer L. Sheppard² and Philip Lavretsky¹, (1)University of Texas at El Paso, El Paso, TX, (2)Simax Ecology, Tauranga, New Zealand

The New Zealand (NZ) Grey Duck (*Anas superciliosa superciliosa*) has come into secondary contact with the closely related Mallard (*A. platyrhynchos*) due to the release of more than 25,000 farm-raised Mallards throughout NZ since 1860. Previous research suggested wide-spread hybridization has resulted in only ~5% of pure Grey Ducks remaining throughout NZ and that the population is now a hybrid swarm. Here, we use a landscape level approach to determine whether hybridization between Mallards and Grey Ducks has resulted in a hybrid swarm, and to identify any barriers that limit hybridization. We collected samples from 673 Mallards, Grey Ducks, and putative hybrids from the North ($N = 378$) and South ($N = 295$) Islands of NZ. Using ddRAD-seq techniques to sequence ~3,500 nuclear loci, we report that Grey Ducks are strongly structured from Mallards ($\Phi_{ST} = 0.085$). However, only 5% and 10% of samples were identified as pure Grey Duck or Mallard, respectively. The remaining 85% of samples comprised a hybrid swarm that resulted from backcrossing primarily into Mallards, suggesting that assortative mating may limit further genetic swamping of Grey Ducks. In general, hybrid prevalence was highest on the North Island and east of the Southern Alps in the South Island. In contrast, most pure Grey Ducks were concentrated in more remote montane habitat west of the Southern Alps, indicating that the Alps act as a geographic barrier to hybridization. Finally, Mallards were genetically differentiated between Islands, suggesting that gene flow is limited by Cook Strait.

PE0262: Natural Populations

A Novel Approach to Investigate the Relationship between Genome Structure and Reduced Reproductive Fitness in Critically Endangered Birds

Jana Renee Wold¹, Andrew Digby², David Eccles³, Nicola Hemmings⁴, Roger A Moraga Martinez⁵, Kylie Price⁶, Joanna Roberts⁷ and Tammy Steeves¹, (1)University of Canterbury, Christchurch, New Zealand, (2)Department of Conservation, Wellington, New Zealand, (3)Gringene Bioinformatics, Wellington, New Zealand, (4)Department of Animal & Plant Sciences, University of Sheffield, Sheffield, United Kingdom, (5)Tea Break Bioinformatics Limited, Palmerston North, New Zealand, (6)Hugh Green Cytometry Core Facility, Malaghan Institute, Wellington, New Zealand, (7)FlowJoanna, Palmerston North, New Zealand

Where recovery programmes use intensive management, selecting appropriate individuals for breeding is vital for mitigating the risk of inbreeding depression, or reduced fitness due to mating between close relatives. Current breeding selection tools aim to minimise inbreeding and maximise genome-wide diversity, but rarely account for genes underlying negative fitness traits, in large part because these are generally unknown. Emerging evidence indicates that genes associated with fitness can be found in chromosome rearrangements, as in male zebra finch (*Taeniopygia guttata*) where sperm morphology is determined by genes caught in an inversion on the Z-chromosome. Despite implications for conservation management, the prevalence of similar rearrangements in other birds is unknown. Determining this is of particular interest for highly inbred bird species, particularly if some breeding individuals have Z-chromosome inversions that harbour disadvantageous alleles at genes associated with

male and female fitness. Here, we present a novel approach for investigating genomic structure using kākāpō (*Strigops habroptilus*) as our Proof of Concept. We will then apply these learnings to intensively managed critically endangered birds in Aotearoa New Zealand, with applicability to similar species-at-risk around the globe.

PO0263: Natural Populations

Genetic Diversity and Footprints of Selection in Wild House Mice

Raman Akinyanju Lawal and Beth Dumont, The Jackson Laboratory, Bar Harbor, ME

A major goal of evolutionary genetics is to decipher the mechanisms of adaptation to diverse environments. Here, we used 154 whole-genome sequences from wild-caught house mice to perform genome-wide scans for selection. These samples span three core house mice subspecies (*Mus musculus domesticus*, *M. m. castaneus*, *M. m. musculus*) and the outgroup taxon, *M. spretus*, and include organisms inhabiting diverse ecological environments. We employ multiple established population genetic methods and conservative statistical thresholds to spotlight 71 loci in *M. m. domesticus*, 52 in *M. m. castaneus*, 48 in *M. m. musculus*, and 50 in *M. spretus* that are evolving non-neutrally. Included among these loci are well-established targets of positive selection among mammals, including olfactory receptors, genes involved in reproduction, and *Epas1*, which has been previously implicated in physiological adaptation to hypoxic environments. Several loci are also evolving via different evolutionary selection regime in different subspecies, including *Hbb-bh2*, *Pr12c3*, *Cntnap2*, *Lrrc25*, and *Susd6*. Our analyses also underscore a key role for balancing selection in the maintenance of genetic diversity at several genes, including *Cwc22*, *Zswim2* and *Fam171b*. Taken together, our findings comprise a catalog of putative signals of positive and balancing selection in a powerful biomedical model system poised for facile discovery of the genetic and underlying physiological mechanisms of adaptation in the wild.

PE0264: Natural Populations

Patterns of Adaptive Genetic Variation across Coffea Canephora

Valérie Poncet, IRD UMR DIADE, Montpellier cedex 5, France

Understanding how organisms respond to their environment by altering physiological processes will increase our capacity to make predictions about adaptation to global climate change. Adaptive clines have been increasingly studied in plant species within temperate zones to understand adaptation of organism in natural populations. However, they are still poorly understood in tropical environments. *Coffea canephora*, cultivated as Robusta, is an interesting tropical tree model to investigate adaptation in the tropics, as it is largely distributed within the range of the lowland tropical rain forests of Africa. In particular, modifications occurring in genes related to abiotic stress tolerance make these genes candidate for enhanced resilience to future climate change. We combined the use of both captured regions sequenced for a set of candidate genes related to drought tolerance and whole genome SNP markers. Leveraging on a robust statistical approach combining multiple neutrality statistics, we provided a comprehensive map of selection signals in the genome of the *C. canephora* both at the species level and within its major genetic groups.

The genotype-environment association suggests regional adaptation to spatially varying environments of the recent past, with a special focus on the Eastern edge of the distribution, in Uganda. More specifically, we found signals of selection tightly linked to several genes involved in response to biotic and abiotic stress and in caffeine biosynthesis. Our detection of selection signals support the hypothesis of present ecological gradient contributing to the structure of the genetic diversity. Moreover, assessing the genomic vulnerability of the present populations will help to predict their response to future environmental changes.

Denoeud, F. *et al.* (2014) The coffee genome provides insight into convergent evolution of caffeine biosynthesis. *Science* **345**, 1181-1184.

Dereeper, A. *et al.* (2015) The coffee genome hub: a resource for coffee genomes. *Nucleic Acids Research* **43**, D1028-D1035.

Marraccini, P. *et al.* (2012) Differentially expressed genes and proteins upon drought acclimation in tolerant and sensitive genotypes of *Coffea canephora*. *Journal of Experimental Botany* **63**, 4191-4212.

Merot-L'anthoene, V. *et al.* (2019) Development and evaluation of a genome-wide Coffee 8.5 K SNP array and its application for high-density genetic mapping and for investigating the origin of *Coffea arabica* L. *Plant Biotechnology Journal* **17**, 1418-1430.

Gomez, C. *et al.* (2016) Shift in precipitation regime promotes interspecific hybridization of introduced *Coffea* species. *Ecology and Evolution* **6**, 3240-3255.

PO0265: Natural Populations

Dynamic Effects of Interacting Genes underlying Rice Flowering-Time Phenotypic Plasticity and Global Adaptation

Tingting Guo¹, Qi Mu¹, Jinyu Wang¹, Adam E. Vanous¹, Akio Onogi², Hiroyoshi Iwata³, Xianran Li¹ and Jianming Yu¹, (1)Iowa State University, Ames, IA, (2)National Agriculture and Food Research Organization, Japan, (3)The University of Tokyo, Tokyo, Japan

Observed phenotypic variation in living organisms is shaped by genomes, environment, and their interactions. The rapidly accumulating genome sequence data advance the identification of novel alleles of important genes underlying phenotypic variation. However, how alleles and combinations of alleles change effects across environments, resulting in phenotypic plasticity, is often unknown. Here, we demonstrate an analytical framework to dissect the phenotypic plasticity and to associate plasticity with geographic haplotype distribution. First, we conducted phenotypic plasticity dissection in a rice genetic population. The observed flowering-time plasticity can be systemically dissected, modeled, and predicted by four key genes (*Hd1*, *Hd2*, *Hd5*, and *Hd6*). These genes, discovered for their photoperiodic response, differentially responded to temperature at the early growth stage to jointly determine flowering time. The effects of these plasticity genes were revealed with multiple reaction norms along the temperature gradient. With the integration of genomics and the temperature environmental index, accurate performance predictions were obtained. Second, we screened the accessions from the 3,000 Rice Genomes Project for allelic variation on the four flowering-time genes and constructed haplotypes at both individual-gene and multi-gene levels. The geographic distribution of haplotypes showed preferential haplotype combinations adapted into different temperature zones. The temperate zone was dominated by haplotypes more sensitive to temperature change, while the tropical zone had a majority of haplotypes less sensitive. Our findings bridged the gap between phenotypic plasticity dissection and allele mining by integrating knowledge from genomics, cloning and function characterization, environment quantification, and predictive modeling.

PE0266: Canine

Next Generation Genotyping (NGG) of *Canis familiaris* Using Igenomx Riptide™ DNA Library Preparation

Paul Doran¹, Azeem Siddique¹, Gaia Suckow², Nils Homer³, Phil Ordhoukhanian², Steve Head², Keith Brown¹, Matt J. Huentelman⁴, Joe Pickrell⁵ and Alvaro G. Hernandez⁶, (1)iGenomX, Carlsbad, CA, (2)The Scripps Research Institute, La Jolla, CA, (3)Fulcrum Genomics, Somerville, MA, (4)Translational Genomics Research Institute, Phoenix, AZ, (5)Gencove, New York, NY, (6)University of Illinois at Urbana-Champaign, Urbana, IL

Dogs have been living with humans for approximately 15000 years. Selective breeding has created a multitude of dog breeds with distinct characteristics. Great interest exists in understanding how selection has affected the modern dog genome and what variants are linked to specific canine breed characteristics. Dogs are also susceptible to a number of diseases that have counterparts in humans. Their unique population structure, relatively limited heterogeneity within breeds, greater genome sequence identity to humans than mice, and their sharing of a common environment with humans make them an excellent model organism for certain human diseases.

The iGenomX RipTide library prep is a high throughput DNA library prep for next generation sequencing that has been used to prepare libraries for a variety of applications where large numbers of samples require library preparation at low cost. One such application is Next Generation Genotyping (NGG). Here we show the use of the RipTide library prep in a case control GWAS study, generating over 30 million biallelic SNPs per sample on a cohort of West Highland White Terriers. After filtering, more than 5.2 million SNPs were identified with a minor allele frequency of >5%. PCA analysis showed that the variants permitted the accurate identification of breeds. The data also showed a novel genetic association with Westie lung disease, the canine equivalent of chronic obstructive pulmonary disease in humans.

The iGenomX RipTide library prep combined with Illumina sequencing generated more variants in less time and at lower cost than the standard microarray-based genotyping experiment.

PO0267: Canine

Functional Annotation of the Dog Genome Using Histone Modifications ChIP-Seq

Jocelyn Plassais¹, Erika M. Kim², Jacquelyn M. Evans¹, Dustin T. Whitaker¹, Heidi G. Parker¹ and Elaine A. Ostrander¹, (1)National Human Genome Research Institute, National Institutes of Health, Bethesda, MD, (2)Cancer Informatics Branch, National Cancer Institute, National Institutes of Health, Bethesda, MD

Domestic dog breeds are characterized by an unrivaled diversity of morphologic traits and breed-associated behaviors, as well as disease susceptibilities, resulting from selective pressures by early humans as well as breeding practices used to develop and maintain modern breeds. Recently, a catalog of 722 canine whole genome sequences (WGS) was published, documenting over 91 million single nucleotide and small indels. While coding mutations can now be more easily identified, it is still difficult to identify variants in non-coding regions, including those in enhancers and promoters. In this context, we have developed an epigenetic approach, focusing on chromatin histone modifications (H3K4me1, H3K4me3, H3K27ac) to identify regulatory elements in the canine genome. Using our veterinarian network, we have sampled eight tissues (liver, spleen, heart, bladder lining, stomach lining, lymph node, bicep femoris, ear cartilage) from four dogs, and also utilized three canine cancer cell lines. We have generated data and, using direct comparison between biological replicates, we have already identified thousands of enhancer and promoter elements within the dog genome. We have also included in our dataset publicly available transposase-accessible chromatin sequencing data (ATAQ-Seq), which identifies open chromatin regions. We thus have generated the largest catalog of genomic regulatory elements for the dog genome community, allowing us to identify candidate variants associated with breed-specific morphological traits.

PE0268: Canine

Composite Selection Signals Application in Purebred Dogs

Victor Wei Tse Hsu, University of Sydney, Darlingtown, NSW, Australia, Mehar S Khatkar, Sydney School of Veterinary Science, The University of Sydney, and ARC Research Hub for Advanced Prawn Breeding, Camden, Australia and Peter Williamson, The University of Sydney, Sydney, Australia

Composite selection signals (CSS) have demonstrated that the locus or interesting regions can be localized in multi-breed populations. The CSS method combines three signals based on population structure, selected or derived allele frequency, and extended haplotype homozygosity (ancestral and convergent sweeps). Here, we investigated the application of CSS to canine SNP data to assess previously known signatures or identify novel regions from various purebred dogs breed comparisons. In addition, we tested the use of CSS to reveal regions associated with canine disease, using lymphoma susceptibility as an example. Multiple studies have identified a predisposition to lymphoma dogs belonging to the Mastiff clade, including Bullmastiffs. To gain a more comprehensive insight into lymphoma predisposition, we performed a population-based study using the CSS method to analyze selected regions for potential impact on lymphoma incidence. 364 Bullmastiffs were used as a target group with a number of comparative reference groups derived from single or combined breed data. A SNP dataset of gray wolves, previously reported in a European study, was used as a source of ancestral alleles. Using the ancestral or convergent sweeps, clusters of signatures of selection were detected at 101 regions on nine canine autosomes. A gene ontology and pathway analysis of genes in regions identified by CSS revealed 89 candidate genes with enrichment for

lymphoma-associated ontologies. The most significant signals were related to the regulation of lymphocyte migration. The CSS is a useful tool for cross-species for identifying potential candidate genes under selection.

PO0269: Canine

A Catalog of Polymorphic SINEC_Cf Insertions in the Dog Genome

Jessica D. Choi¹, Sara E. Kalla², Allison Seebald³, Max Tomlinson³, Yun Seok Lee¹, Cassandra R. Ward¹ and Nathan B. Sutter¹, (1)La Sierra University, Riverside, CA, (2)Baylor College of Medicine, Houston, TX, (3)Cornell University, Ithaca, NY

Retrotransposons are ubiquitous in mammal genomes where they can account for nearly half of all sequence. The genome of the domestic dog, *Canis familiaris*, harbors the L1 LINE and several types of SINEs. In particular, the dog reference genome contains 171,386 annotations for SINEC_Cf, a SINE originating from tRNA. Many SINEC_Cf copies are very young, including some that have not yet gone to fixation. To discover polymorphic SINEs we constructed 62 sequencing libraries enriched with sequence flanking SINEC_Cf by hybridizing a primer to well-conserved sequence in SINEC_Cf's head that is usually absent from other SINE types. These libraries represent 59 breeds that broadly survey dog genome variation. We identified 81,747 putative polymorphic SINE insertions detected in at least one of our libraries but absent from the Boxer reference genome. Because SINEs are known to disrupt normal patterns of gene expression and splicing, we identified areas in dog genes where reference genome SINEs or LINEs are excluded or have a strand bias. For example, LINEs and SINEC_Cfs are excluded from introns near exons and both types of retrotransposons have a strand bias for intronic insertions. SINEC_Cf insertions in introns near exons can cause diseases such as narcolepsy; we find many insertions less than 30 bp from exons. Finally, we also find dozens of polymorphic SINE insertions in protein-coding exons (many in known pseudogenes) and hundreds of SINEs in UTRs and promoters. The high insertion rate of SINEC_Cf provides a natural mutagenesis screen in the dog genome.

PE0270: Canine

Characterization of Polymorphic SINE Insertions and Genes in Dog Retrotransposon Free Regions

Yun Seok Lee¹, Sara E. Kalla², Allison Seebald³, Jessica D. Choi¹ and Nathan B. Sutter¹, (1)La Sierra University, Riverside, CA, (2)Baylor College of Medicine, Houston, TX, (3)Cornell University, Ithaca, NY

Retrotransposons are mobile genetic elements that have played a major role in mammalian genome evolution. For example, retrotransposon insertions in the dog genome have introduced novel open reading frames and splice acceptor sites, and caused phenotypes ranging from narcolepsy and other diseases to the merle coat pattern selected within some breeds. One dog retrotransposon in particular, SINEC_Cf, is so young that thousands of insertions have not yet gone to fixation. Despite the presence in the dog reference genome of 1,351,940 LINEs and 1,134,572 SINEs (of which 171,386 are SINEC_Cf), we have identified 1375 "free regions" that are at least 10,000 bp long and contain no SINEs, LINEs, or assembly gaps. There are 16,901 free regions at least 5000 bp long, many of which span over gene upstream or downstream ends. We have analyzed the genes found in these dog SINE+LINE free regions because transposon free regions in the human and mouse genomes were previously shown to be rich in genes crucial for early development and transcriptional regulation. We have also analyzed patterns of polymorphic SINE insertion into our free regions to check whether SINEs in these loci have lower than average insertion frequencies or tend to insert at free region edges. To make this possible we Illumina sequenced 434 libraries created by extending into flanking non-repeat sequence from a primer hybridizing to conserved SINEC_Cf sequence. The libraries represent 356 dogs from 125 breeds.

PO0271: Canine

A Genome Wide Association Study of Performance Traits in Alaskan Sled Dogs

Joseph Thorsrud and Heather J Huson, Cornell University, Ithaca, NY

The objective of this research is to find the genetic basis for performance traits in sled dogs. Alaskan sled dogs are one of the only populations of canine which have been selected based almost entirely on performance. This unique

evolutionary history can lead to insight into the loci associated with traits of interest for sled dogs, as well as other working dogs, and athletes in general. These traits are – work ethic, leadership, heat tolerance, mental and physical stress tolerance, endurance, and speed.

SNPs on chromosomes 2, 5, 6, 8, 9, 11, 15, 17, 20, 22, 24, 26, 27, 33, and 37 have been identified as associated with speed. Chromosomes 5, 14, 25, and 29 contained SNPs associated with work ethic. Endurance was associated with SNPs on chromosome 2. Heat tolerance was associated with SNPs on chromosome 34. There were SNPs associated with physical stress tolerance on chromosomes 3, 16, and 22. Mental stress tolerance was associated with SNPs on chromosomes 7 and 12.

Now that the loci associated with these traits have been found, the ancestry of these traits can be identified since sled dogs represent an admixture of other purebred dogs. Specific gene functions that are related to the loci can also be identified to better understand the biochemical pathways that are being utilized in these traits. With this knowledge, breeders can select for dogs that carry advantageous alleles in not only sled dogs, but military, search and rescue, racing, and police dogs as well.

PE0272: Canine

Integrating Genome-Wide Association Study and Animal Insurance Data to Identify Genetic Region Associated with Skin Problem in Dogs

Yuki Matsumoto¹, Napat Ruamrungsri² and Genki Ishihara¹, (1)Anicom Specialty Medical Institute Inc., Kanagawaken, Japan, (2)Anicom Specialty Medical Institute Inc., Yokohamashi-Nakaku, Japan

Long term medical data sets such as insurance is poorly available in the field of animal genetics, although such data could be contributed to uncover genetic bases related to disease. Here we collected insurance data derived from 119 client-owned Miniature Dachshund insured by a Japanese insurance company, Anicom Insurance, Inc. during the period between 2008–2019 as well as buccal swab for extracting DNA of the dogs. We used a single nucleotide polymorphism (SNP) array, CanineHD 230K Whole-Genome Genotyping BeadChip for genotyping. PLINK software was used for quality control for SNPs and for Genome-wide association studies with free of skin problem. We found a locus associated with skin problem in Chromosome 10. A candidate gene located in the locus which expressed in the human skin illustrates the gene could affect skin problem.

PO0273: Canine

Automated DNA and RNA Purification from Canine Feces, Serum and Buccal Swabs Using the Maxwell® RSC Instrument

Jennifer Mook, Leta Steffen and Brad Hook, Promega Corporation, Madison, WI

Molecular based laboratory testing and research is providing specialized pathways to help veterinarians diagnose and treat diseases. Specific methodologies are needed for purifying nucleic acid from different biological samples types as each sample source provides its own benefits and challenges. Feces is an abundant sample source for the study of the host microbiome but contains inhibitors that can interfere with PCR and next generation sequencing (NGS). The Maxwell® RSC Instrument was used to purify DNA from canine feces and it was assessed for inhibition using qPCR with 16S bacterial rRNA gene primers. Canine serum is also a common sample in veterinary laboratories and has the potential to be a source for biomarkers. Serum sample collection is invasive and of limiting supply, so maximizing yield is important especially for downstream applications such as NGS. MicroRNAs were purified from canine serum using the Maxwell® RSC and assessed using 2-step RT-qPCR for miRNAs 26b and 29a. Molecular based veterinary testing is also utilized for genetic testing of breed-specific genetic mutations. Buccal swabs are a cost-effective, non-invasive means of collecting canine DNA. The Maxwell® RSC was used to purify DNA from canine buccal swabs and assessed by qPCR. The Maxwell® RSC instrument is a versatile magnetic bead platform that allows for user-friendly workflows that can handle both simple and challenging sample types for veterinary research applications.

PE0274: Canine

The Genetic Underpinnings of Histiocytic Sarcoma in a Canine Disease Model

Jacquelyn M. Evans, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD

Dogs are a naturally-occurring model for many complex human disorders, including cancer. Histiocytic sarcoma (HS) is a rare hematological malignancy in humans with a histologically similar counterpart, occurring at high frequency in the Flat-coated Retriever (FCR) breed. To identify the genetic basis of this breed predisposition, a genome-wide association study was performed in 177 cases diagnosed by histopathology or cytology, and 132 aged, controls with no cancer history, using the 170k Illumina Canine HD SNP array. A strong signal was identified on chromosome 5, showing extensive regional linkage disequilibrium. Analysis of shared haplotypes among cases delineates a conservative 1.2Mb critical interval. Similarly, B-cell lymphoma in Golden Retrievers (GRs) has been associated with this chr5 locus. FCRs and GRs are closely-related breeds, sharing a recent common ancestor, and the two cancers derive from cells in the hematopoietic stem cell pathway. Taken together, these similarities suggest there may be overlapping disease mechanisms. Indeed, a ~300kb shared risk haplotype was identified among cases of both breeds. Whole genome resequencing data of affected and healthy FCRs has been generated to identify putative candidate causal variants for HS, with the data pointing to regulatory regions, which have previously been largely unexplored in canine disease models.

PO0275: Canine

Phenotypic Classification and Genetic Variability of Canine Neutrophil Subpopulations

Elizabeth Pavlova, Cornell University, Ithaca, NY

Neutrophils, an essential cell of the innate immune system, destroy invading bacteria through various mechanisms, including phagocytosis. Once considered a homogenous population of cells, accumulating data indicate neutrophils are phenotypically diverse based upon differences in density (high-density neutrophils, HDN, and low-density neutrophils, LDN) and cell surface markers between subpopulations of neutrophils. However, a paucity of canine research in this area leaves these neutrophil subpopulations mostly undefined. This project aims to explore and characterize different subpopulations of neutrophils in blood samples obtained from a cohort of retired sled dogs part of a more extensive study investigating canine aging. A relationship between reduced longevity and more copies of the retroelement, Long Interspersed Nuclear Element-1 (LINE-1), in the genome of dogs predicates this broader study. This study aims to observe and classify three subpopulations of neutrophils that were found to have different phagocytic activity in preliminary assays; to optimize markers and functional assays to characterize those subpopulations further; to conduct a Genome-Wide Association Study (GWAS) on the sled dog cohort to explore connections between genotype and the relative sizes of neutrophil populations and illness phenotypes, especially cancer; and to compare proportions of neutrophils in these populations with LINE-1 copies, fully integrating this project into the broader goal of the canine aging study. Blood will be obtained through venipuncture and neutrophils will be isolated by density gradient separation. Markers will be added to these isolated cells and then the cells will be plated with latex beads and incubated for several hours to observe phagocytosis, then run on a flow cytometer to measure the levels of phagocytosis. The results of flow cytometry will show different cell populations on forward and side scatter, which will be gated and assessed as different neutrophil subpopulations based on differences in phenotype in phagocytosis. A positive correlation is expected between LINE-1 copies and the number of LDN in circulation, consistent with prior research done on myeloid-derived suppressor cells (MDSC), a type of LDN associated with canine neoplasia.

PE0276: Canine

Complete Workflow Solutions for High- and Low-Density Canine Genotyping from Thermo Fisher Scientific

Tao Lei, Thermo Fisher Scientific, Santa Clara, CA

Genotyping applications can be varied based on customer needs and requirements. High to low density genomic testing can be useful, whether it is for a veterinary disease research company or a canine breeding program. Applied Biosystems™ Axiom™ Canine HD Array, a high-density genotyping platform, is useful for genome wide association studies and marker assisted selection for traits and diseases. Applied Biosystems™ AgriSeq™ targeted GBS platform, a low density targeted GBS method which allows for marker assisted selection for traits and diseases,

identification of novel mutations, parentage verification and identification of individuals based on genotypes. Having one complete workflow solution from sample prep to genotyping for both platforms allows for more convenience and versatility.

Here, we demonstrate a single extraction used for various genotyping platforms. Using the Applied Biosystems™ MagMAX CORE AgGenomic Module, extractions from canine buccal swabs were tested on two different genotyping platforms. AgriSeq™ Canine SNP Parentage and ID Panel, was combined with AgriSeq™ Canine Traits and Disorders Panel, to target 535 total markers on the AgriSeq™ targeted GBS platform. The same extracted samples were also tested on the Axiom™ Canine HD Array, which provides over 710,000 markers. The data supports a complete solution from DNA extractions to genotyping for companion animals and livestock from Thermo Fisher Scientific.

PO0277: Canine

Detection of Genome-Wide Copy Number Variation Using the Applied Biosystems™ Axiom™ Genotyping Solution

Ali Pirani, Ram Varma and Mohini A. Patil, Thermo Fisher Scientific, Santa Clara, CA

Copy number variations (CNVs) have been implicated in both disease phenotypes and phenotypic variation associated with quantitative traits in plant and animal species. Software tools available from Thermo Fisher Scientific offer a valuable resource for evaluation of the impact of CNVs with respect to variation in plant and animal health and production traits. Applied Biosystems™ Axiom™ Analysis Suite Software now has enhanced CNV capabilities that utilize intensity and genotypes to calculate log₂ ratios and B-allele frequencies (BAFs) from genotyping data for detailed analysis and visualization.

This platform has the dual ability to detect CNVs in targeted genomic regions and to genotype single nucleotide polymorphisms (SNPs) across the whole genome using a single assay. CNV analysis methods include (a) whole genome de novo analysis for discovery and (b) fixed region analysis when breakpoints of CNV regions of interest are known a priori and there is little breakpoint variability from sample to sample. In CNV discovery analysis, CN states are determined by a Hidden Markov Model implementation. Breakpoints are discovered and CN segments are labeled. Fixed region analysis uses an optimized multi-sample clustering algorithm to assign CN states to each region in each sample.

These software tools allow the user to perform complex copy number analysis utilizing both methods. This results in superior analytical sensitivity and specificity for known small regions, while enabling discovery for larger regions and across the whole genome. Examples of copy number gains and losses are presented here.

For Research Use Only. Not for use in diagnostic procedures.

PE0278: Feline

The Genomic History of Tigers from the Pleistocene to Present - an Ancient DNA Perspective

Xin Sun, The State Key Laboratory of Protein and Plant Gene Research, Peking-Tsinghua Center for Life Sciences, School of Life Sciences, Peking University, Beijing, China; Academy for Advanced Interdisciplinary Studies, Peking University, Beijing, China and Shu-Jin Luo, The State Key Laboratory of Protein and Plant Gene Research, School of Life Sciences, Peking-Tsinghua Center for Life Sciences, Peking University, Beijing, China

Three of the nine tiger subspecies have gone extinct during the 20th century leaving range-wide wild tigers at less than 4,000 today and *Panthera tigris* one of the most charismatic conservation icons. The oldest tiger fossil was dated at 2-3 Mya, however mitochondrial coalescence of modern tigers was traced to only 110 kya, suggesting a Late Pleistocene bottleneck and complex demographic dynamics, which can only be elucidated through analysis of the extinct tigers. Here, we sequenced the first ancient tiger genome dated at ~10,600 BP uncovered from the Russian Far East (RFE), in addition to museum specimens representing the extinct tigers from the wild, including 12 South China tigers (*P. t. amoyensis*) and three Caspian tigers (*P. t. virgata*). Combined with published genomes from 32 voucher tigers, phylogenomic analysis supported *P. t. amoyensis* as a statistically robust clade albeit its mitochondrial paraphyly, thus resolving the long-lasting taxonomic controversy. The ancient RFE tiger carried a basal mitochondrial lineage distinct from modern Amur tigers (*P. t. altaica*), however clustered in the autosomal tree

within the northeast Asian phylogroups including *P. t. amoyensis* and *P. t. altaica*. At last, *P. t. virgata* of central Asia originated from tigers in the RFE expanding westbound and subsequently had genetic introgression from Bengal tigers to the south. Ecological niche modeling (ENM) showed that climate dynamics during the Last Glacial Period has dramatically influenced the habitat suitability for tigers, likely leading to multiple range contraction-expansion-isolation cycles and consequent phylogeographic distinction of living tiger subspecies to date.

PE0280: Feline

Chromosome-Level Assembly for an American Shorthair Cat Genome

Yuki Matsumoto¹, Kei Watanabe¹, Genki Ishihara¹, Hideki Hirakawa², Sachiko Isobe², Ting-Fung Chan³, Hon-Ming Lam³, Yasuhiro Tanizawa⁴ and **Yasukazu Nakamura⁴**, (1)Anicom Specialty Medical Institute Inc., Kanagawaken, Japan, (2)Kazusa DNA Research Institute, Kisarazu, Chiba, Japan, (3)Center for Soybean Research of the SKL of Agrobiotechnology, and School of Life Sciences, The Chinese University of Hong Kong, Hong Kong, Hong Kong, (4)National Institute of Genetics, Mishima, Japan

We introduce our complete genome project which aims to provide a high-quality assembly of a female domestic cat of American Shorthair breed for the research community. We will describe the detailed strategy and the current progress of the project. Also, we propose and discuss on a system of precision medicine for companion animals with use of a high-quality reference genome and massive multi-omics information, which we collect from many numbers of patient cats. Reference genome sequence with detailed annotation and raw-data from NGS's obtained in this project will be opened without any restriction for the research community through the INSDC (DDBJ/ENA/NCBI) SRA and Assembly databases, as well as our own www-based database. As methodology, we employ following five technologies to construct a set of pseudomolecule for each chromosome of the cat: (1) deep sequencing with use of Pacbio sequencer, (2) error correction of the assembly by illumina reads, (3) chromosome-level scaffolding process by the chromatin conformation capture sequencing (Hi-C), (4) optical mapping by the Bionano Saphyr and (5) Iso-Seq by Pacbio for better gene-structure annotation.

PO0281: Feline

Morphology and Genetics of Kinked Tails of Domestic Cats (*Felis catus*) in East and Southeast Asia

Hao Meng¹, Xiao Xu¹, Worata Klinsawat², Pramana Yuda³ and Shu-Jin Luo¹, (1)The State Key Laboratory of Protein and Plant Gene Research, School of Life Sciences, Peking-Tsinghua Center for Life Sciences, Peking University, Beijing, China, (2)School of Bioresources and Technology, King Mongkut's University of Technology Thonburi, Bangkok, Thailand, (3)Fakultas Teknobiologi, Universitas Atma Jaya Yogyakarta, Yogyakarta, Indonesia

The form of tail plays a critical role in animal survival and adaptation. A considerable proportion of domestic cats in Southeast Asia exhibit characteristic short/kinked tails that may have persisted for centuries whose underlying evolutionary mechanisms, however, remain unknown. Mutations in coding sequences (CDS) of *T* and *HES7* have been identified correlated to the short tails in Manx and Japanese bobtail breeds as well as some feral cats in Asia. However, about one third of short-tailed cats in China do not carry either variant, suggesting the existence of at least a third gene causative of the tail length polymorphism in Felidae. We conducted morphological surveys of domestic cats in China (N=617), Thailand (N=380) and Indonesia (N=382), and revealed a gradually increased prevalence of short/kinked tails from the north (5.8%) to the south (71.2%), consistent with a putative Southeast Asian origin of the trait. No indel in CDS of *T* was detected and the *HES7* mutation was only found in less than 40% of kinked-tailed cats we sampled (N=340). Further, a pedigree including 36 cats was built to segregate the unknown short-tail-causal gene, and a Mendelian autosomal dominant mode of inheritance was confirmed. Whole genome sequences were assembled from 29 cats with an average 10X genome coverage. GWAS in unrelated cats and linkage analysis in the pedigree were conducted to identify a 1.6-Mb genomic region associated with the trait. These findings promise to help understand the evolutionary mechanism and adaptive roles of mammalian tail diversity.

PE0282: Feline

Endogenous Retrovirus Insertions in *KIT* and Their Associations with White Spotting in Domestic Cats

Yueting Xing, Xiao Xu, Hao Meng, He Yu and Shu-Jin Luo, The State Key Laboratory of Protein and Plant Gene Research, School of Life Sciences, Peking-Tsinghua Center for Life Sciences, Peking University, Beijing, China

Dominant white and white spotting (the *W* locus) are among the most common “domestic” traits and prevalent in numerous domestic species. Naturally occurring “white gloving” has also been reported in free-ranging wildlife. In domestic cats, the coat color patterns of dominant white (*W* allele) and white spotting (*w^s* allele) are known to be caused by the insertion of a feline endogenous retrovirus (FERV1), at various lengths, into *KIT* the Felidae *W* locus (*w⁺* refers to the wild-type allele without white spotting). Here we further examined domestic cats with various level of white patches, aiming at elucidating the underlying fine-scale phenotypic effects of associated *w^s* alleles. Samples from 790 cats were collected, 421 of which with full-body images were scored for individual degrees of white spotting. Worldwide allele frequency survey showed that dominant white and white spotting were the most prevalent in China and Asia respectively. We also detected variations in the FERV1 insertions representing different *w^s* alleles. Among *w^s* carriers, those with two *w^s* alleles present higher degree of white spotting relative to *w^s/w⁺*. In *w^s/w⁺* cats, we identified a *w^s* allele corresponding to less white portion, compared to other *w^s* alleles. All *w^s* and *W* haplotypes formed a clade in the phylogenetic tree based on 47.8-kb sequences surrounding the *KIT*, suggesting a single origin of white color during cat domestication and that *W* alleles were derived from *w^s*. These findings expand our understanding of the genetics of white spotting in Felidae and the evolutionary impact of feline endogenous retroviruses.

PO0283: Feline

Comprehensive Panel-Based Mutation Screening of Feline Disease Variants in 6000 Domestic Cats

Heidi Anderson¹, Katherine M. Lytle², Oliver P Forman³, Stephen Davison³, Leena Honkanen¹, Julia Mathlin¹, Rebecca Chodroff Foran², Hannes Lohi⁴ and Jonas Donner¹, (1)Wisdom Health, Helsinki, Finland, (2)Wisdom Health, Vancouver, WA, (3)Wisdom Health, Leicestershire, United Kingdom, (4)University of Helsinki, Helsinki, Finland

Introduction: There are over 80 cat breeds, types, or varieties recognized by different cat registries today. The majority of these breeds have been developed within the past 50-70 years. Random-bred cat populations have provided the foundation for all cat breeds, and continue to be exploited for outcrosses and as foundation stock for new breeds.

Objective: Our objective for this study was to use a comprehensive panel-based mutation screening test to generate up-to-date information about feline disease variant distribution and prevalence across breeds. Such information has important veterinary diagnostic and breeding implications.

Materials & Methods: A total of 6003 felines – 5612 pedigreed cats and 391 random-bred cats – were genotyped for the presence of 45 Mendelian disease variants previously described in domestic cats. This large pedigreed cat study population represents 86 breeds, types, and varieties recognized by at least one major international cat registry: Fédération Internationale Féline (FiFe), The International Cat Association (TICA), The Cat Fanciers' Association (CFA), and World Cat Federation (WCF). Most cats were from North America (49.3%) and Europe (48.1%), with cats from New Zealand and Australia (1.2%) representing another notable subgroup. Genotyping was carried out on a custom-designed Illumina® Infinium HD genotyping array commercially available as the MYCATDNA™ / OPTIMAL SELECTION™ Feline Genetic Breeding Analysis test.

Results: Of the 45 tested feline disease variants, 22 (48.9%) were observed at least one time in this study cohort. While 13.4% of the tested cats were heterozygous and 0.7% of the cats were homozygous for at least one of the disease variants, the maximum number of the disease variants found in any individual cat was three. There were 18 disease variants observed in pedigreed cats, and six of these mutations were also found in random-bred cats. Four disease mutations were solely observed in the random-bred cat population. The maximum number of disease variants seen in a single breed was six. The most common feline disease variants present in the study sample were Pyruvate Kinase Deficiency (4.9% of cats) and Progressive Retinal Atrophy (*rdAc*) (3.1% of cats).

Conclusions: This study represents the most comprehensive exploration of feline disease heritage to date in terms of Mendelian disease variant coverage across virtually all known cat breeds. Some disease variants were widely distributed across cat populations, underscoring the value of panel screening for disease variants as a cost efficient process with immediate practical relevance for breeding selection and veterinary care.

PE0284: Feline

Spatiotemporal Genetic Diversity of Lions

Caitlin Curry¹, Brian W. Davis¹, Laura Bertola², Paula A. White³, William J. Murphy¹ and James N. Derr¹, (1)Texas A&M University, College Station, TX, (2)City College of New York, (3)Zambia Lion Project, University of California, Los Angeles, Los Angeles, CA

We determined the genetic architecture of both historical and modern lions to identify changes in genetic diversity that occurred over 100 years of landscape and anthropogenic change. We surveyed microsatellite and mitochondrial genetic variation from 143 high-quality museum specimens of known provenance and combined them with data from recently published nuclear and mitochondrial studies. Analysis of variation at 9 microsatellites and 280 polymorphic mitogenome SNPs indicate the presence of male-mediated gene flow and recent isolation of local subpopulations, likely due to habitat fragmentation. Nuclear markers showed a significant decrease in genetic diversity from the historical ($HE=0.833$) to the modern ($HE=0.796$) populations, while mitochondrial genetic diversity was maintained ($Hd=0.98$ for both). While the historical population appears to have been panmictic based on nDNA data, hierarchical structure analysis identified four tiers of fine structure in modern populations, able to detect most sampling locations. Mitochondrial analyses identified 4 clusters: Southern, Mixed, Eastern, and Western; and were consistent between modern and historically sampled haplotypes. Within the last century, habitat fragmentation caused lion subpopulations to become more isolated as human expansion changed the African landscape. This resulted in an increase in fine-scale nuclear genetic structure and loss of genetic diversity as subpopulations became more differentiated, while mitochondrial structure and diversity was maintained over time.

PO0285: Equine

Assessing the Impact of Sequencing Platform on Transcriptome Assembly, Differential Expression, and Variant Discovery in the Horse

Jonah N. Cullen¹, Robert J Schaefer², Sian Ann Durward-Akhurst¹, James R Mickelson¹ and Molly E McCue¹, (1)University of Minnesota, St. Paul, MN, (2)University of Minnesota, St Paul, MN

The purpose of this project was to examine the impact of next generation sequencing platform on the analysis of both RNA-seq and whole genome sequencing (WGS) data from equine samples. RNA-seq libraries were prepped from muscle and fat tissues and sequenced on both the Illumina HiSeq and NovaSeq systems. RNA-seq analysis was conducted using a publicly available containerized workflow developed by our research group as part of a larger effort to build an equine tissue expression atlas. This workflow employs both STAR/StringTie and Salmon to enumerate gene and transcript expression thus enabling quantification-specific comparisons between commonly used methods. WGS data was generated from buffy coat samples and processed using a modified version of the GATK variant discovery best practices workflow, also developed by our research group. In addition to the expected platform-specific differences in output (*e.g.* total reads per run), we examined analysis-associated disparities in mapping rates and accuracy, transcriptome assembly and completeness, differential expression, tissue-specific gene co-expression networks, and discrepancies in variant discovery. Furthermore, we compared the normalized transcript expression matrices produced by STAR/StringTie or Salmon within each sequencing platform. This project will enhance our understanding of how sequencing platform may affect critical downstream output including transcriptome completeness, differential expression, and variant discovery in the horse. Additionally, the results provided herein will serve to better inform how data collected as part of the community-wide tissue expression atlas project should be considered and merged.

PE0286: Equine

Whole Genome Imputation in the Horse

Robert J Schaefer¹, Samantha K Beeson², Sian Ann Durward-Akhurst², Bart A Grantham³, William Knox Carey⁴, James R Mickelson² and Molly E McCue², (1)University of Minnesota, St Paul, MN, (2)University of Minnesota, St. Paul, MN, (3)Interval Bio, San Mateo, CA, (4)Interval Bio, Mountain View, CA

The MNEc670k SNP genotyping array was constructed to allow researchers to reliably genotype up to 670,805 SNPs throughout the equine genome. Variants present on the array were selected based on several criteria, including the ability to reliably tag haplotype windows among 24 different horse breeds, allowing for efficient genotype imputation. The process of genotype imputation allows for a target set of SNPs to be computationally inferred to a larger SNP set based on the haplotypes present in a reference population, as well as filling in missing or incorrect genotypes. The success and accuracy of imputation depends on the SNP density as well as the cohort size of the reference population, from which the haplotypes are inferred.

During the design of the MNEc670k SNP chip we demonstrated imputation up to ~2M SNPs with 96-99% accuracy in a reference population containing 332 horses. Similarly, using a larger SNP panel (~16M SNPs) discovered from whole genome sequence (WGS) in Standardbreds, our lab has shown 89% imputation accuracy in a cohort of 179 horses. To date, the biggest hurdle in reliably performing genome level imputation in the horse has been a fragmented WGS variant calling pipeline, mapping to an obsolete genome assembly (EquCab2), as well as the lack of joint genotype calls leading to ambiguity between missing and homozygous reference genotypes.

To better enable whole genome imputation in the domestic horse, we have processed variant calls generated from whole genome sequences in a cohort of 549 horses mapped to EquCab3 using a standardized pipeline implemented in collaboration with Interval Bio. Over 32 million variants were detected using joint genotype calling from both GATK as well as bcftools. Using this cohort, we present imputation accuracies based on scenarios among SNP sets from previous generations of genotyping arrays (SNP50, SNP70, MNEc670k and MNEc2M) as well to whole genome levels. Additionally, we show differences in imputation accuracy based on the breed specific reference populations as well as other imputation parameters.

PO0287: Equine

Functionally Annotating Regulatory Elements in the Equine Genome Using Histone Mark ChIP-Seq

Nicole B. Kingsley, University of California, Davis, Davis, CA, Colin Kern, University of Delaware, gsggf, AL, Catherine Creppe, Diagenode, Liège, Belgium, Erin N. Burns, University of California - Davis, Davis, CA, Huaijun Zhou, Department of Animal Science, University of California, Davis, CA, Ted Kalbfleisch, University of Louisville, Louisville, KY, James N. MacLeod, Department of Veterinary Science, University of Kentucky, Lexington, KY, Jessica L. Petersen, University of Nebraska-Lincoln, Lincoln, NE, Carrie J. Finno, UC Davis, Davis, CA and Rebecca R. Bellone, Veterinary Genetics Laboratory, University of California - Davis, Davis, CA

As part of the Functional Annotation of ANimal Genomes (FAANG) initiative, we aimed to characterize tissue-specific regulation within the equine genome. Chromatin Immunoprecipitation Sequencing (ChIP-Seq) was used to map four histone modifications (H3K4me1, H3K4me3, H3K27ac, and H3K27me3) in eight prioritized tissues (adipose, brain, heart, lamina, liver, lung, muscle, and ovary) collected from two adult Thoroughbred mares. Data were generated according to optimized experimental parameters developed during quality control testing. To ensure sufficient IP and successful peak-calling, data and peak-calls were assessed using six quality metrics, replicate comparisons, and site-specific evaluations. Additionally, combinations of marks were used to identify active regions unique to each tissue. We found that liver and lamina were the tissues with the most unique active elements (19,028 and 17,206, respectively) while brain and liver had the most tissue-specific repressed regions (1,610 and 1,606, respectively). Tissue specificity was further explored by identifying binding motifs within the unique active regions, and motifs were characterized by gene ontology (GO) term and protein-protein interaction network analyses. The histone marks identified in this study represent some of the first publicly available resources for investigating tissue-specific regulation within the equine genome. As such, these annotation data can be used to advance equine studies investigating health, performance, reproduction, and other traits of economic interest in the horse.

PE0288: Equine

The Frequency of Loss of Function Alleles in the Equine Population

Sian Ann Durward-Akhurst¹, Robert J Schaefer², James R Mickelson¹ and Molly E McCue¹,
(1)University of Minnesota, St. Paul, MN, (2)University of Minnesota, St Paul, MN

Identification of disease-causing alleles is a fundamental goal of medical genetics and can facilitate disease diagnosis and improve disease understanding. Studies in humans have increasingly demonstrated the value of databases of genetic variation derived from genome sequencing (WGS) for disease-causing allele(s) identification. A surprising finding from these studies is the high number of alleles computationally predicted to lead to loss of function (LOF) of the affected gene in healthy adults. LOF variants with a high frequency in healthy individuals are unlikely to be disease-causing. Here, we investigate LOF alleles present in the equine population.

We mapped WGS from 549 horses to the EquCab3 reference genome. Single nucleotide polymorphisms (SNPs) and short insertions/deletions were identified using GATK-HaplotypeCaller and SAMtools, and annotated using ANNOVAR and SnpEff. The intersect was used to identify LOF variants i.e., nonsense, frameshift, and splice site disrupting variants, or deletions removing the first exon or >50% of protein coding sequence. Average depth of coverage was 10.3x (range 1.4 - 37.4x). 32,825,825 variants were identified, with 15,067 predicted to have high impact including 7,575 LOF variants affecting 4,126 genes enriched for olfactory reception and immune related pathways. On average, each horse carried 1,333 (range 387 - 1,869).

Overall, we demonstrate that similar to humans, LOF alleles are present in the horse population. We will further validate the LOF alleles using hand annotation and produce a list of LOF alleles that can be excluded from candidate disease-causing allele discovery approaches due to their high frequency in the general population.

PO0289: Equine

Equine Y Chromosome Variability

Caitlin Castaneda¹, Andrew Hillhouse², Sabine Felkel³, Barbara Wallner³ and Terje Raudsepp¹,
(1)Texas A&M University, College Station, TX, (2)Department of Veterinary Pathobiology, Texas A&M University College of Veterinary Medicine, College Station, TX, (3)Institute of Animal Breeding and Genetics, Vetmeduni Vienna, Vienna, Austria

Sequence copy number (CN) and copy number variation (CNV) cause phenotypic variations, however consequences of CN gain or loss in horses remains enigmatic despite the updated horse genome EquCab3 and recent completion of the horse Y chromosome reference assembly. Here, we continue CN and CNV studies of Y-linked ampliconic genes between individuals and breeds, and between reproductively normal males, subfertile males and horses with disorders of sexual development (DSD). We have generated reproducible data by ddPCR for 8 ampliconic genes and the single-copy *SRY* gene across 20 selected breeds which include Thoroughbred-related commercial breeds, Przewalski's horse, Oriental and isolated breeds. We also include a cohort of 10 subfertile males with spermatogenic defects and 32 horses of ambiguous sex. We observe high degree of CN conservation across Thoroughbred related breeds with Testis-Specific Protein on Y (*TSPY*) having the greatest range (7-17 copies) and Equine Testis Specific Transcript Y 2 (*ETSTY2*) having low variability between normal individuals (3-4 copies). Interestingly, CN of *ETSTY* genes vary more between subfertile and normal males, suggesting that CN stability of ampliconic genes may be functionally important. Additionally, CN analysis of *SRY* in horses with DSD shows that a loss of *SRY* is accompanied by CN loss in neighboring ampliconic genes. A novel finding is increased CN of *SRY* accompanied with CN changes (loss or gain) of neighboring ampliconic genes in phenotypically normal individuals of isolated indigenous breeds and Przewalski's Horses. We are currently expanding CN research to compare Y CNVs with SNP and haplotype variations using the same cohort. The findings are expected to reveal comparative contribution of CNVs and SNPs to Y chromosome variation in horses, to gain a deeper understanding of horse Y variability.

PE0290: Equine

Integrating Transcriptomics, Proteomics, and cis-Regulatory Networks to understand Muscle Physiology and Pathophysiology

Deborah Velez-Irizarry, Kennedy J. Aldrich, Zoë J. Williams and Stephanie J. Valberg, Department of Large Animal Clinical Sciences, Michigan State University, East Lansing, MI

Gene set enrichment analysis (GSEA) is standard practice for large scale omics experiments. GSEA uses a priori gene lists from curated databases to group experimentally derived gene sets into biological pathways in order to infer underlying physiological states. A transcription factor motif enrichment analysis (TFMEA) follows the same basic principles as GSEA except inferences are made on cis-regulatory states and functional transcription factor (TF) binding. When applied to multi-omics datasets, TFMEA and GSEA can recognize coordinated physiological and pathophysiological processes by identifying cis-regulatory networks in differentially expressed gene sets. We have applied this integrated approach using transcriptomics and proteomics to study myofibrillar myopathy (MFM, n=16), recurrent exertional rhabdomyolysis (RER, n=15) and glycogen depletion/repletion (GDR, n=40) in horses. We first identified differentially expressed (DE) genes/proteins for each project using weighted linear regression/permutation tests ($FDR \leq 0.05$) followed by TFMEA ($NES \leq 3.0$) and GSEA ($FDR \leq 0.05$) for putative target genes (PTG). We focused on TFs differentially regulated in disease or metabolic state. This approach identified enriched TF DE (eTF-DE) in horses with MFM ($\uparrow ATF3$, an anti-inflammatory factor) in association with upregulation of its PTG involved in protein phosphorylation. In RER, eTF-DE ($\uparrow HSF4$) identified PTG associated with the mitochondrial TCA cycle and nitrogen metabolism; *HSF4* is also a potential target of the RER treatment dantrolene. eTF-DE in response to GDR identified $\uparrow RUNX1$ with PTG enriched for G protein-coupled receptor downstream signaling involved in glucose homeostasis. Our research shows the added value of TFMEA for functional genomic studies of equine skeletal muscle disorders and glycogen metabolism.

PO0291: Equine

A Likelihood Estimate Method for Detecting Introgressed Alleles from Non-Caballine Equids in Horses

Kalpani de Silva¹, Ernest Bailey² and **Ted Kalbfleisch**², (1)University of Louisville, Louisville, KY, (2)University of Kentucky, Lexington, KY

The horse has two alleles of *CXCL16*, one associated with susceptibility and one with resistance to developing persistent shedding of the Equine Arteritis Virus. The two alleles differ by 4 non-synonymous variants in exon 1 of the gene. Comparison with 3 non-caballine equids (zebras, asses and hemionids) revealed that one haplotype was almost identical to the haplotype found in non-caballines while the other had differences characteristic of 4.5 million years since a common ancestor. Based on this observation, we project that an ancient introgression event occurred between caballine and non-caballine equids. If so, we should be able to find more instances of introgression between these species. We developed a method to identify putatively introgressed segments in the horse genome. It is estimated that non-caballine equids such as zebras and asses diverged from horses between 4 and 4.5 MYA. Genomic analysis of these animals vs. EquCab3 reveals the divergence at both the nucleotide and chromosomal level. Whole genome data for the non-caballine equids show a greater frequency of single nucleotide differences than horses have relative to one another. We have created a Likelihood Estimate framework that uses this difference in single nucleotide frequencies to predict whether a haplotype evolved along the caballine or non-caballine lineage. Preliminary data shows these haplotypes to be between 0.5 and 2kb in length and are detectable at a rate of several hundred loci per horse. Furthermore, these haplotypes occur at high frequency in the horse population suggesting that they are beneficial alleles, and perhaps evidence of adaptive introgression.

PE0292: Equine

Comparison of Poly-A+ Selection and rRNA Depletion in Detection of lncRNA in Two Equine Tissues Using RNA-Seq

Anna R. Dahlgren¹, Erica Scott², Tamer A. Mansour³, Erin N. Burns¹, Pablo J. Ross⁴, Ted Kalbfleisch⁵, James N. MacLeod⁶, Jessica L. Petersen⁷, Rebecca R. Bellone⁸ and Carrie J. Finno⁹, (1)University of California - Davis, Davis, CA, (2)School of Veterinary Medicine, UC Davis, Davis, CA, (3)University of California, Davis, CA, (4)Department of Animal Science, University of California, Davis, CA, (5)University of Kentucky, Lexington, KY, (6)Department of Veterinary Science, University of Kentucky, Lexington, KY, (7)University of Nebraska-Lincoln, Lincoln, NE, (8)Veterinary Genetics Laboratory, University of California - Davis, Davis, CA, (9)UC Davis, Davis, CA

Long non-coding RNAs (lncRNAs) are untranslated regulatory transcripts longer than 200 nucleotides (nt) that can play a role in transcription, post-translational modification, and epigenetic regulation. Traditionally, RNA-sequencing (RNA-seq) libraries have been created by isolating transcriptomic RNA via poly-A⁺ selection. In the past 10 years, methods to perform ribosomal RNA (rRNA) depletion of total RNA have been developed as an alternative, with the goal of allowing for better coverage of whole transcriptomic RNA, including both polyadenylated and non-polyadenylated transcripts. The purpose of this study was to determine which library preparation method is optimal for investigation of lncRNAs in the horse. Using liver and parietal lobe of the cerebrum tissues from two healthy Thoroughbred mares, RNA-seq libraries were prepared using standard poly-A⁺ selection and rRNA-depletion methods. Poly-A⁺ selection yielded 617 and 4939 more lncRNA transcripts for liver and parietal cortex, respectively. lncRNA expression was similar between biologic replicates for both library preparation methods when mapped to EquCab2 based on principle component analysis. However, mapping to EquCab3 showed that poly-A⁺ selection had more consistency in lncRNA expression between the two horses compared to rRNA-depletion. On the other hand, rRNA-depletion was better able to detect >200nt long small nucleolar RNAs (snoRNAs) as annotated by the National Center for Biotechnology Information or Ensembl. Overall, poly-A⁺ selection provides a more thorough identification of total lncRNA in horses while rRNA-depletion may allow for more accurate detection of snoRNA.

PO0293: Equine

Genome Scan for Back and Croup Conformation in the Icelandic Horse

Maria K Rosengren¹, **Marina Solé**¹, Åsa Viklund¹, Susanne Eriksson¹, Brandon D Velie², Juan J Negro³ and Gabriella Lindgren^{1,4}, (1)Department of Animal Breeding & Genetics, Swedish University of Agricultural Science, Uppsala, Sweden, (2)Faculty of Life and Environmental Science, University of Sydney, Sydney, Australia, (3)Department of Evolutionary Ecology, Doñana Biological Station, CSIC, Sevilla, Spain, (4)Livestock Genetics, Department of Biosystems, KU Leuven, Leuven, Belgium

A flexible and well-muscled top line (back and croup) of adequate length and shape is considered important for Icelandic horses and included in the breeding goal. The trait is moderately heritable and favorably genetically correlated with competition performance. Despite moderate heritability for most conformation traits, only a few single nucleotide polymorphisms (SNPs) have previously been reported to be associated with one other conformation trait (neck, withers and shoulders) in the Icelandic horse. The aim of this study was to identify genomic regions associated with the trait back and croup conformation in Icelandic horses by using linear scores from breeding field tests on a scale from 5 to 10. One hundred seventy-eight (44% males and 56% females) horses genotyped using the 670K+ Axiom Equine Genotyping Array were included in the study. A genome-wide association study was performed considering a mixed model-structured association approach with the mmscore function in the package GenABEL in R, followed by an haplotype analysis with the haplo.stats package in R. Quality control (QC) included call rate > 0.90, missing genotypes in individuals > 0.90, minor allele frequency < 0.05 and cut-off p-value 1e-10 for Hardy-Weinberg Equilibrium. After QC, 387924 autosomal SNPs and 177 horses remained for further analysis. Twelve SNPs on Equus Caballus (ECA) 22 reached the suggestive threshold (P-value < 1.0 x 10⁻⁵). Haplotype analysis revealed two opposite haplotypes which resulted in higher and lower scores for the back and croup (P-value < 0.001). The genomic region harbors genes associated with anthropometric traits in humans.

PE0294: Equine

The Effects of Inbreeding on Covering Success, Gestation Length, and Foal Sex Ratio in Thoroughbreds

Evelyn T. Todd, University of Sydney, Sydney, NSW, Australia, Natasha A. Hamilton, Racing Australia Equine Genetics Research Centre, Scone, NSW, Australia, **Brandon D. Velie**, Faculty of Life and

Environmental Science, University of Sydney, Sydney, NSW, Australia and Peter C Thomson, School of Life and Environmental Sciences, The University of Sydney, and ARC Research Hub for Advanced Prawn Breeding, Camden, Australia

Horses have an eleven-month gestation period, both conceiving and giving birth in the spring. This long gestation period and short breeding season make the maintenance of good fertility rates in horse populations imperative to provide commercial returns for domestic breeds. Inbreeding can increase the frequency of deleterious variants, thus resulting in reduced reproductive levels. In this study, we examined the influence of inbreeding levels on covering success, gestation length and secondary sex ratio in Australian Thoroughbred mares. Phenotypic data were obtained from 27,262 breeding records of Thoroughbred mares provided by three Australian stud farms with inbreeding estimated based on pedigrees dating back to the foundation of the breed. While both gestation length and covering success were heritable, no measurable effect of inbreeding on either trait was found. However, the genetic value for both traits decreased within recent generations and a number of environmental factors had significant effects on covering success and gestation length. The lack of measureable effects of inbreeding in this study may be due to the level of inbreeding in the sample not being high enough to show a discernible effect on reproductive traits. It is possible that the intensive management techniques used in the Thoroughbred population may also mask any negative effects of inbreeding. Further monitoring of these traits in future generations would assist in understanding the selective forces influencing these traits.

PO0295: Equine

Kinetics of Gene Expression Profiles in Divergent Chondrogenic Pathways of Equine Interzone and Anagen Cell Cultures

Chan Hee Mok, Rashmi Dubey, Emma N. Adam and James N. MacLeod, Department of Veterinary Science, University of Kentucky, Lexington, KY

In fetal limb buds, interzone and anagen cells progress through two different chondrogenic pathways to develop articular cartilage and anagen cartilage, respectively. While articular cartilage remains stable in synovial joints, anagen cartilage progresses through hypertrophic differentiation to form bones. Thus, understanding the comparative cell biology between interzone and anagen cells during development may provide novel insights into emergent cell-based therapies to support articular cartilage regeneration. To assess the kinetics of gene expression profiles after inducing chondrogenesis in culture, primary cells from seven equine fetuses were generated and compared at ten time points (0, 1.5, 3, 6, 12, 24, 48, 96, 168, and 336 hrs). Total RNA was isolated, and gene expression on a targeted set of 93 gene loci was measured in a microfluidic RT-qPCR system. Differential transcriptional responses were observed as early as 1.5 hr after the initiation of chondrogenesis. Genes with functional annotations that include transcription regulation responded to the chondrogenic induction earlier (1.5 – 96 hr) than genes involved in signaling cascades (1.5 – 336 hr) and the extracellular matrix (3 – 336 hr). Between interzone and anagen cell cultures, expression levels of 73 genes were not initially different at 0 hr, but 47 out of the 73 genes became differentially expressed under the chondrogenic stimulation. Studies with these targeted gene loci, as well as full transcriptomic analyses, are being used in an effort to identify key regulator(s) responsible for the divergent developmental pathways of interzone and anagen cells.

PE0296: Equine

Mapping CTCF Binding Regions in the Equine Genome

Sichong Peng¹, Catherine Creppe², Nicole B. Kingsley¹, Erin N. Burns³, Rebecca R. Bellone⁴, Jessica L. Petersen⁵ and Carrie Finno⁶, (1)University of California, Davis, Davis, CA, (2)Diagenode, Liège, Belgium, (3)University of California - Davis, Davis, CA, (4)Veterinary Genetics Laboratory, University of California - Davis, Davis, CA, (5)University of Nebraska-Lincoln, Lincoln, NE, (6)University of California Davis, Davis, CA

CCCTC-binding factor (CTCF) is a zinc finger protein that serves as a core architect protein in chromosome 3D structures, which dictates the chromosome regulatory activities. CTCF binds at chromosomal domain boundaries and often functions as an insulator, blocking inter-domain DNA interactions. Chromatin immunoprecipitation (ChIP) can be used to pull down CTCF-bound DNA and subsequent sequencing of enriched DNA can identify these bound regions. As part of Functional

Annotation of Animal Genomes (FAANG) project, we utilized ChIP-seq to sequence CTCF bound regions on eight prioritized tissues (adipose, brain, heart, lamina, liver, lung, muscle, and ovary) from two healthy adult Thoroughbred mares. The specificity and efficiency of four different anti-CTCF antibodies were evaluated through qPCR using primer sets for both positive (H19 imprinting control region) and negative (myoglobin, exon 2) regions. Libraries were combined into four pools and each pool sequenced using Illumina HiSeq 3000/4000. An average of 28M reads were obtained from each library. Reads were mapped to EquCab3 reference genome after QC and peaks were called using MACS2 narrow peak setting for each library. An average of 53,000 peaks were identified for each tissue type after merging biological replicates (adipose: 51k, brain: 49k, heart: 63k, lamina: 53k, liver: 46k, lung: 52k, muscle: 48k, and ovary: 68k). ChIP-seq for histone marks from the same tissue samples were then incorporated with CTCF data and ChromHMM was used to predict chromatin states in eight prioritized tissues.

PO0297: Equine

Gene Expression Profiling of the Pelvic Flexure and Surrounding Large Colon in the Equine Gastrointestinal Tract By RNA-Sequencing

Amber Lynne Wilson, Kaysie Jennings and Stephen J. Coleman, Colorado State University, Fort Collins, CO

The equine large intestine is essential both for nutrient absorption and as a location for microbial digestion. Previously, we identified the pelvic flexure as an important anatomic and physiologic landmark dividing distinct microbial communities – several taxonomic groups were present in the cecum and ventral colon that were not present in dorsal or small colons. Since no physical barrier separates these compartments, we became curious about other mechanisms that could segregate the different microbial groups into these different areas. The physiology of each compartment could support different interactions with resident bacteria and by extension a distinct microbiota at each location. The objective was to evaluate gene expression in the pelvic flexure and surrounding large colon to characterize the physiological processes active in these tissues. Samples of mucosal and submucosal layers of the right and left ventral colon, the right and left dorsal colon, and the pelvic flexure were collected from three American Quarter Horses immediately postmortem and RNA isolated using TRIzol™. Single-end libraries were prepared using the NEBNext® Ultra II Directional RNA library kit and sequenced on an Illumina® NextSeq platform. Sequencing generated an average of 23 million reads per sample. Gene expression analysis was performed by aligning quality trimmed sequence reads to EquCab3 using the HISAT2 and StringTie algorithms. Assembled transcripts were compared to publicly available annotations. Investigation of gene expression in this gastrointestinal region will help enhance our understanding of digestive physiology and support future investigations of the processes which regulate interactions between host and microbiota.

PE0298: Equine

Phenotyping Approaches for Identification of Heritable Variation in Startle Response

Barclay B. Powell, Carissa L Wickens, Emily K Miller-Cushon, Andreas Keil and Samantha A. Brooks, University of Florida, Gainesville, FL

The startle response, a reflexive behavioral reaction to the sudden appearance of a novel stimulus, influences the survival of animals in the wild. In humans, excessive startle response is a clinically significant endophenotype of a number of psychiatric conditions. For domestic animals, inappropriate startle response behaviors can injure the animal and/or human handler. In livestock species like the horse, heritable components of the startle response are important genetic trait for breeding and management.

We describe a longitudinal study observing startle response and fear behavior in stock-type horses. Young horses raised as part of the UF Equine educational program were assessed as weanlings (n=72, mean age=256 days) and/or as two-year olds (n=72, mean age=762.4 days). A total of 96 horses participated in the study, comprising five different foal crops over four consecutive years. A behavioral assay based on presentation of a sudden novel stimulus (rapidly opened umbrella) was used to trigger and document the startle response. Physiological measurements recorded throughout the trial included heart rate (HR), latency to return to the site of the novel object (seconds), maximum distance fled (meters), and other additionally noted post-startle behaviors quantified by an

ethogram such as defecation and touching the umbrella after re-approaching. Correlating genomic relatedness of the observed horses with the standardized phenotype data produces a heritability estimate for aspects of the startle response, a possible genetic trait suitable for study. Using the Genome-wide Complex Trait Analysis package (GCTA). In particular, we examined the trends between an individual horse's weanling and two-year-old test results for heart rate change just after startle. The measure of startle reflexes changing with maturity may indicate variation in retention of neotenus behaviors, a common feature of domesticated species. The genomic heritability for variation in startle response with age is at 39.8%.

Ongoing research is focused on working with collaborators to further analyze phenotypes, such as heart rate acceleration or decelerations during the startle response, to better quantify this complex behavior. Our goals in the near future are to produce a reproducible and accurate measure suitable for future genome wide association study to identify novel loci influencing the startle trait. Our long-term goal is to continue expanding the sample size through additional future foal crops to validate our behavioral findings and improve statistical power.

PO0299: Equine

Single Nucleotide Polymorphism Characterization of Major Histocompatibility Complex Haplotype Diversity in Standardbred Horses

Lucia Younger¹, Annette M McCoy², Donald Miller¹, Molly E McCue³, James R Mickelson³ and Douglas Antczak¹, (1)Cornell University, Ithaca, NY, (2)University of Illinois, Urbana, IL, (3)University of Minnesota, St. Paul, MN

Identification of Major Histocompatibility Complex (MHC) haplotypes remains challenging because of the expected high, but unknown, level of diversity within any given species and the difficulty of determining polymorphism across the approximately 4 million base pair extent of the MHC. In the horse, the MHC has been characterized using various methods including serology and microsatellites. Here we explored MHC haplotype diversity in the equine Standardbred breed using 670K SNP chip data from 297 Standardbred horses. Single Nucleotide Polymorphisms (SNP) in the MHC region were extracted and analyzed using the bioinformatic software SHAPEIT. This allowed identification of haplotype-specific SNP patterns and phasing of MHC SNP-based haplotypes in MHC heterozygous horses. We found a total of 80 unique haplotypes comprised of 47 class I and 20 class II blocks. Twenty-five MHC homozygous horses were identified that each carried only one of six common haplotypes. Intra-MHC microsatellite testing is underway to determine the relationship between SNP-based and microsatellite-based MHC haplotypes. Thus far, tests performed on DNA from 31 of the horses identified 17 microsatellite haplotypes, 8 of which had previously been described in Standardbreds, and complete correlation between SNP and microsatellite haplotypes. SNP-based MHC typing holds promise for characterization of polymorphism of the equine Major Histocompatibility Complex. Understanding the amount of diversity in the MHC has many applications in respect to the evolution and physiology of the horse.

PE0300: Equine

Genome-Wide Association Analyses of Equine Metabolic Syndrome Phenotypes in Welsh Ponies and Morgan Horses

Elaine M Norton¹, NE Schultz¹, Ray Geor², Dianne McFarlane³, James R Mickelson¹ and Molly E McCue¹, (1)University of Minnesota, St. Paul, MN, (2)Massey University, Palmerston North, New Zealand, (3)Oklahoma State University, Stillwater, OK

Equine metabolic syndrome (EMS) is a complex trait for which few genetic studies have been published. The study objectives were to perform within breed genome-wide association analyses (GWA) to identify associated loci in two high-risk breeds, coupled with meta-analysis to identify shared and unique loci between breeds. GWA for twelve EMS traits identified 303 and 142 associated genomic regions in 264 Welsh ponies and 286 Morgan horses, respectively. Meta-analysis demonstrated that 65 GWA regions were shared across breeds. Region boundaries were defined based on a fixed-size or the breakdown of linkage disequilibrium, and prioritized if they were: shared between breeds or across traits (high priority), identified in a single GWA cohort (medium priority), or shared across traits with no SNPs reaching genome-wide significance (low priority), resulting in 56 high, 26 medium, and 7 low priority regions including 1,853 candidate genes in the Welsh ponies; and 39 high, 8 medium, and 9 low priority

regions including 1,167 candidate genes in the Morgans. The prioritized regions contained protein-coding genes which were functionally enriched for pathways associated with inflammation, glucose metabolism, or lipid metabolism. These data demonstrate that EMS is a polygenic trait with breed-specific risk alleles as well as those shared across breeds.

PO0301: Equine

Metabo-Genomics of Equine Metabolic Syndrome – An Analogous Condition to Human Metabolic Syndrome

Laura Patterson Rosa, Martha F. Mallicote, Maureen T. Long and Samantha A. Brooks, University of Florida, Gainesville, FL

Equine metabolic syndrome (EMS) is an analogous condition to human metabolic syndrome leading to type 2 diabetes (T2D). Although obesity, hyperinsulinemia, and peripheral insulin resistance (IR) are diagnostic hallmarks of EMS, affected horses may also exhibit hypertension, hyperlipemia and systemic inflammation. Measures of IR typically comprise the gold-standard for diagnosis in veterinary care. Yet, the dynamic nature of insulin homeostasis and complex procedures required for typical assays make accurate quantification of IR in the field challenging. Furthermore, IR may also result from underlying disease processes, being a direct effect rather than a cause of the condition.

Examination of the biochemistry at play through global metabolomics can identify useful biomarkers for early detection of disease. In humans, a targeted approach revealed specific compounds significantly altered during the pre-diabetic state. Additionally, two metabolites proved to be predictive for development of T2D as early as seven years prior to diagnosis. Yet, identification of individual metabolites can be challenging. Former analysis of combined genomics and global metabolomics has provided novel knowledge of the origin and pathways of several diseases, including T2D.

We applied a global liquid chromatography/mass spectroscopy approach (HPLC/MS) to whole plasma collected from a pilot population of 49 Arabian horses, resulting in 3392 high-confidence features. A composite score derived from 9 common diagnostic variables quantified EMS disease in the 49 horses. A linear regression model with the log-transformed values for each metabolite feature identified 10 significant compounds ($p \leq 1.474e-5$) correlated with the EMS score. For each metabolite, we estimated heritability (H^2) with genotypes from the 670k Affymetrix Equine SNPchip for each of the study animals. Seven of ten metabolite features are highly heritable ($H^2 = 0.59-0.99$) and are excellent targets for future genome-wide association studies in larger populations. The identified genomic loci will provide insight in to the pathways controlling variation in these metabolites, and the origin of genetic predisposition to EMS. Rapid, feasible and accurate diagnostic tools can be translated into measurable benefits in the timeline and quality of preventative management practices to preserve the health in these horses.

PE0302: Equine

Whole Genome Sequencing Identifies Missense Mutation in GRM6 As the Putative Cause of Congenital Stationary Night Blindness in a Tennessee Walking Horse

Yael Leah Hack, University of California- Davis, Davis, CA, Elizabeth Crabtree, North Carolina State University College of Veterinary Medicine, Raleigh, NC, Felipe F Avila, UC Davis, Davis, CA, Roger B. Sutton, Texas Tech University Health Sciences Center School of Medicine, Lubbock, TX, Brian C. Gilger, North Carolina State University, Raleigh, NC and Rebecca R. Bellone, Veterinary Genetics Laboratory, University of California - Davis, Davis, CA

Congenital stationary night blindness (CSNB) is an inherited, non-progressive retinal disorder characterized by absence of vision under low-light conditions. Currently, the only known genetic cause of CSNB in the horse is an insertion in *transient receptor potential cation channel subfamily M member 1* (*TRPM1*, ECA1g.

108,297,929_108,297,930 ins1378.) However, one Tennessee Walking Horse diagnosed with CSNB did not have this mutation. To identify a causal variant, Illumina Novaseq whole genome sequence data from this case was compared to data from horses from seven other breeds ($n=33$). One hundred and two candidate genes, identified from human and mouse literature, were assessed for coding variants. Variants in these candidate genes homozygous in the case and absent in all other horses were prioritized for further investigation. A single missense mutation in *metabotropic glutamate receptor 6* (GRM6) (c.533C>T p.Thr178Met), a gene known to cause CSNB in humans,

was identified. This SNP was predicted to be deleterious with 61% confidence by the consensus classifier PredictSNP. Thr178 is highly conserved across vertebrate species and is directly involved in binding the neurotransmitter glutamate, and is thus essential to on-bipolar cell signaling that enables vision in low light conditions. Methionine at position 178 is hypothesized to impair binding of glutamate, which was supported by protein modeling. In screening 80 unrelated Tennessee Walking Horses, the estimated allele frequency was 8.1% with no other homozygotes identified. Taken together, these data provide evidence that this SNP is causal for CSNB in this breed. Additional testing is warranted to evaluate if this variant is present in other breeds with CSNB.

PO0303: Equine

Differential Gene Expression Provides Translational Evidence for the Basis of Equine Myofibrillar Myopathy

Zoë J. Williams¹, Deborah Velez-Irizarry¹, Jessica L. Petersen², Carrie J. Finno³ and **Stephanie J. Valberg⁴**, (1)Department of Large Animal Clinical Sciences, Michigan State University, East Lansing, MI, (2)University of Nebraska-Lincoln, Lincoln, NE, (3)UC Davis, Davis, CA, (4)Michigan State University, East Lansing, MI

Myofibrillar myopathy (MFM) disrupts the skeletal muscle sarcomere and in humans is associated with mutations in 16 genes with 50% of cases having an unknown and potentially complex etiology. MFM in Warmblood (WB) horses shares histopathologic features of human MFM but a familial basis has not been established. Our hypothesis was that variants in the 16 genes associated with human MFM would *not* segregate with WB MFM, and further, that transcriptomic analysis and enriched pathways would elucidate pathomechanisms causing MFM. mRNA from the *gluteus medius* of 8 MFM WB and 8 control WB was sequenced (Illumina HiSeq 4000). Aligned to EquCab3.0, variants were called in candidate genes and allele frequencies compared among MFM, control, and variant data from whole-genome sequence of 33 non-phenotyped WB using a Fisher's exact test and Bonferroni correction ($P < 0.05$). Across the transcriptome, differential expression (DE) analysis was performed using a negative binomial generalized log-linear model and multiple test correction ($FDR < 0.05$). There were 26 missense variants found in 11 of the candidate genes, but none significantly associated with MFM WB. There were 44 DE genes in MFM *vs.* control WB with log₂ fold changes between -6 and 4.8. Gene Ontology analysis found enrichment in *negative regulation of protein and cellular metabolism* (GO:0051248, GO:0031324) and *positive regulation of reactive oxygen species metabolism* (GO:2000379) in MFM WB. Although MFM in WBs was not associated with candidate genes, the myofibrillar disarray seen in MFM WB may be due to oxidative stress and its associated impacts on metabolism and protein turnover.

PE0304: Equine

Inherited Hypocalcemia in Thoroughbred Foals Is Associated with a Nonsense Mutation in RAPGEF5

Victor Rivas¹, Gary Magdesian², Sophia Fagan³, Mustafa Khokha³ and Carrie J. Finno⁴, (1)University of California-Davis, Davis, CA, (2)University of California Davis, Davis, CA, (3)Yale University, New Haven, CT, (4)UC Davis, Davis, CA

Inherited Hypocalcemia (IH) is a neonatal disease that occurs only in Thoroughbreds. The disease is characterized by a profound hypocalcemia with inappropriately normal parathyroid hormone (PTH) concentrations, loss of maneuverability of limbs, seizures, muscle fasciculations, ileus, tachycardia, synchronous diaphragmatic flutter, and ataxia. Previously, DNA from two IH-affected foals, their dams, and two unrelated clinically healthy Thoroughbred horses underwent next-generation sequencing and a whole-genome association study was performed. A segregating nonsense variant in exon 28 of *RAPGEF5* was significantly associated with the IH phenotype. In 2019, samples from an additional two suspect IH foals were collected and Sanger sequencing revealed that both foals were homozygous for the nonsense variant. Preliminary screening suggests that this variant is present in low frequency in the Thoroughbred population. Out of 81 genotyped Thoroughbred horses, only three were carriers for the variant ($q = 0.019$). In order to test if the IH equine *RAPGEF5* mutation affects protein function, we overexpressed the wildtype and IH variant protein in frog embryos. Overexpression of *RAPGEF5* produced a stereotypical phenotype that was not observed in the IH variant. This assay suggested loss-of-function of the protein with the IH equine

mutation. Therefore, there is strong evidence that the nonsense variant in *RAPGEF5* is associated with the IH phenotype.

PO0305: Equine

Differential Gene Expression Analysis Reveals Pathways Important in Early Post-Traumatic Osteoarthritis in an Equine Model

Annette M McCoy¹, Ann M. Kemper¹ and Troy N Trumble², (1)University of Illinois, Urbana, IL, (2)University of Minnesota, St. Paul, MN

Post-traumatic osteoarthritis (PTOA) is common and its progression cannot be reversed by currently available drugs. The ability to identify disease onset and predict progression are priorities in improving the management of PTOA. Our objective was to quantify gene expression changes in the synovium of OA-affected joints in a novel large animal model specifically designed to recapitulate early PTOA. We hypothesized that early PTOA-related changes in the joint are reflected by altered synovial gene expression, particularly in genes that fall within pathways related to inflammation and cellular response to stimuli.

All work was performed under IACUC approval. Briefly, an osteochondral fragment is created in one metacarpophalangeal joint (MCPJ) at the proximal dorsomedial aspect of the first phalanx, then replaced in the fragment bed. The opposite MCPJ is sham-operated. After a 2 week recovery period, the horses are treadmill-exercised for 14 weeks after which the fragment is removed. Synovial samples were collected arthroscopically from the MCPJ of 11 adult horses before (pre-OA) and after (OA) experimental induction of OA, and from sham-operated joints. After sequencing of synovial RNA, quality control was performed with fastq. Salmon was used to quasi-map reads to EquCab3.0 and quantify transcript abundances. Tximport was used to quantify genes for downstream analysis. TMM normalization (edgeR package) was followed by surrogate variable analysis (SVA package). Differential expression (DE) analysis was performed with the limma-treat method using a fold-change cutoff of 1.1. Functional annotation was performed with EnrichR at FDR < 0.05.

RNA was successfully extracted from 28 samples (6 preOA, 11 OA, 11 sham). Sequencing yielded 15.7-29.4 million paired-end reads per sample. 'Sham' and 'preOA' samples were not different and were grouped. 321 genes were upregulated and 351 genes were downregulated in OA synovium compared to unaffected. Gene ontology (GO) terms related to extracellular matrix (ECM) organization and growth factor binding were overrepresented among DE genes. There were 20 significantly enriched pathways; these included pathways involved in ECM turnover, O-glycosylation of TSR domain-containing proteins, and growth factor signaling.

Most enriched pathways and overrepresented GO terms in our data reflect a state of high metabolic activity and tissue turnover in OA-affected tissue, suggesting efforts at healing and restoring homeostasis. TSR domain-containing proteins play a role in many processes including inflammation, development, and wound healing. Limitations of this study include a small sample size and capture of a single time point post-injury. Additionally, changes in gene expression do not always result in changes in protein expression; work to address this point is ongoing. DE genes falling within key pathways may represent potential diagnostic markers or therapeutic targets for PTOA.

PE0306: Equine

Phenotypic and Genetic Characterisation of Recurrent Exertional Rhabdomyolysis in Warmbloods and Connemara Ponies.

Victoria Lindsay¹, Kathleen Selhorst¹, Emily Clark², Richard J Piercy¹ and **Androniki Psifidi^{1,2}**, (1)Royal Veterinary College, University of London, Hatfield, United Kingdom, (2)The Roslin Institute and R(D)SVS, University of Edinburgh, Edinburgh, United Kingdom

Exertional rhabdomyolysis (ER) is a syndrome characterised by episodes of exercise-induced muscle stiffness, contractures, and myofibre necrosis. The disease is clinically heterogeneous and histologically is characterised by non-specific features of muscle damage. Recurrent ER (RER) is a heritable disease in horses ($h^2=0.41-0.49$),

however the understanding of its aetiopathogenesis is limited. RER is typically associated with Thoroughbred racehorses, but a variety of other breeds are diagnosed with RER. This study aims to identify specific phenotypic patterns and investigate the genetic architecture of RER in Warmbloods and Connemaras.

Principal component analysis (PCA) of 205 RER cases was carried out using reported clinical signs, biopsy histological features and signalment; k-means clustering used to assign clusters; features selected using Python sklearn; and χ^2 testing was performed to identify statistical significant differences between clusters with Bonferroni adjustment. Two of the three clusters identified separated based on the stage of muscle damage and regeneration of the biopsy, however one novel RER phenotype was identified, consisting of 11% of the dataset and significantly associated with clinical signs of ataxia ($p=1.48E-4$), gait abnormalities ($p=4.44E-16$), reluctance to go forward ($p=3.17E-9$) and painful muscles ($p=2.67E-6$).

40 horses (10 cases and 10 controls, of each Warmbloods and Connemaras) were whole-genome sequenced (WGS) using the Illumina Hi-SeqX platform. From the WGS data the sequencing of 50 candidate genes based on previous ER studies in humans and horses were extracted. Comparisons between cases and controls within and across breeds were performed to identify variation with a potential functional impact on the encoded proteins.

PO0307: Equine

Comparison of Horses with Juvenile Onset Lordosis to Normal Horses using Whole Genome Sequence

Navid YousefiMashouf, Ted Kalbfleisch and Ernest Bailey, University of Kentucky, Lexington, KY

Juvenile onset lordosis is a conformation defect in horses manifest in the neonate or during the first two years of life and appears as a drop in the topline, often called swayback. Previous work in this laboratory demonstrated a phenotypic prevalence of 5% in American Saddlebred horses. A genome wide association study (GWAS) demonstrated a recessive mode of inheritance and association with a region on horse chromosome 20. The purpose of this study is to identify the genetic variant(s) associated with the trait. For this purpose, whole genome sequence data at ~20-25X of coverage was generated on two lordotic Saddlebreds and a normal-backed Saddlebred with no history of having any lordotic offspring. All the genomic variants from the above three horses were extracted and pooled with WGS variants from six other normal horses for comparison, including Twilight, two other Thoroughbreds, one Standardbred and two horses of unknown breed. In total, 1,161 variants on the genomic region ch20: 41,931,756 – 45,986,467 were found to be homozygous for the non-reference allele in lordotic horses, as compared to being heterozygous or homozygous for the reference allele in the other studied animals. Among them, 621 were within genes, 107 were within exons and 17 were in coding regions. Variant screening efforts are ongoing to discriminate among the candidate variants based on genomic annotations and association verification at the population level.

Keywords: American Saddlebred Horse, Swayback, Lordosis, Whole Genome Sequencing, Fine-mapping

PE0308: Equine

A 16 Kilobase Deletion on ECA13 Is Strongly Associated with Distichiasis in Friesian Horses

Erin A. Hisey¹, Hanneke Hermans², Zachary T. Lounsberry³, Felipe F Avila⁴, Robert A. Grahn⁵, Kelly Knickelbein⁶, Tomasz Ząbek⁷, Ted Kalbfleisch⁸, Mary Lassaline⁹, Willem Back¹⁰ and Rebecca R. Bellone³, (1)University of California, Davis School of Veterinary Medicine, Davis, CA, (2)Utrecht University, Utrecht, Netherlands, (3)Veterinary Genetics Laboratory, University of California - Davis, Davis, CA, (4)UC Davis, Davis, CA, (5)University of California, Davis, Davis, CA, (6)Veterinary Medical Teaching Hospital, University of California-Davis, Davis, CA, (7)Laboratory of Genomics National Research Institute of Animal Production, Balice, Poland, (8)University of Kentucky, Lexington, KY, (9)Mid-Atlantic Equine Medical Center, Ringoes, NJ, (10)Ghent University, Merelbeke, Belgium

Distichiasis, an ocular condition reported in Friesian horses, occurs when aberrant lashes grow from the Meibomian glands of the eyelid. These lashes can cause chronic irritation and corneal ulcers, potentially leading to vision loss

from corneal scarring. Because of its bilateral nature and prevalence in a breed with known inherited monogenic disorders, this condition is hypothesized to be a simply inherited Mendelian trait. To test this hypothesis, a genome wide association study (GWAS) was performed using the Axiom 670k Equine Genotyping array (MNEc670k) on 14 cases and 38 controls clinically phenotyped for distichiasis. With an additive single locus mixed linear model (EMMAX) approach, a 1.83 Mb locus on ECA5 and a 1.34 Mb locus on ECA13 were identified that reached genome-wide significance ($p_{\text{corrected}}=0.016$ and 0.032 , respectively). Only the locus on ECA13 withstood replication testing ($p_{\text{combined}}=3.01 \times 10^{-5}$). A 371-kb run of homozygosity on ECA13 was found in 13 of the 14 cases providing evidence for a recessive mode of inheritance. A haplotype analysis (hapQTL) narrowed the region of association on ECA13 to 163 kb. Whole-genome, high-throughput sequencing data from 3 cases and 2 controls identified a 16-kb deletion from the ECA13 associated haplotype that contains reported putative regulatory elements. This deletion was strongly associated with distichiasis, as 18 of the 19 cases were homozygous ($p=6.0 \times 10^{-10}$). Further functional analyses of this variant will clarify its role in lash development.

PO0309: Equine

Exploration of Chestnut Coat Color Variation via *SALL1*.

Vada E. Hammons and Samantha A. Brooks, University of Florida, Gainesville, FL

Selective breeding in horses has proved to be an invaluable tool in shaping a number of phenotypes to suit human needs. In particular, desirable coat colors can drive large increases in market value of an individual horse, incentivizing research in the field of genetic control of coat color variations. Chestnut coat color in horses is caused by a missense mutation within the *MC1R* gene. However, the intensity of the chestnut color can vary widely within this genotype. The combination of two types of melanin, eumelanin and pheomelanin, create the mixing of black and red pigments, that are visually perceived as a single changing shade. Yet, the genetic explanation for such broad shade variation in chestnut horses has not been investigated. Here, we examined variation in the shade of the chestnut coat color using photographs of ninety-seven horses. Each horse was ranked within the cohort for the shade phenotype by three blinded observers. In genome-wide association study utilizing the relative shade ranking as the phenotype, and five SNPs genotyped using the Affymetrix Equine 670k array, we identified a single associated region on chromosome 3 ($p=2.934 \times 10^{-8}$). Analysis of available whole genome sequences for horses of diverse color using the Integrated Genomics Viewer and UCSC Genome Browser provided a the candidate SNP within the coding sequence of the only gene in the region: *SALL1*. Genotyping of this candidate SNP is ongoing.

The function of *SALL1* is largely unknown. However, protein-protein interactions observed with *HPS1* strengthens the case for its influence on coat color. In humans, a mutation in the *HPS1* coding sequence impairs the development of melanosomes, resulting in partial albinism termed Hermansky-Pudlak syndrome. Thus, altered interaction of *SALL1* and *HPS1* may affect maturation of pheomelanosomes to eumelanosomes, leading to phenotypes with varying levels of pigmentation in the horse. However, with only one study suggesting a circumstantial influence of *SALL1* protein on pigmentation, more research is needed to fully explore the function of this protein. Completion of this study will pinpoint the genetic control of variation within the chestnut coat color, providing information to create new markers to be applied in future selective breeding decisions. Further, this knowledge may reveal potential negative pleiotropic effects of variation in *SALL1*, as previously noted for several coat color loci.

PE0310: Equine

Genome-Wide Association Study of Roan Coat Color in Clydesdales

Alexa Barber¹, Leah Treffer^{1,2}, Renae L. Sieck¹, Isabel Graizan¹, Anna M. Fuller¹ and Jessica L. Petersen¹, (1)University of Nebraska-Lincoln, Lincoln, NE, (2)Nebraska Wesleyan University, Lincoln, NE

The roan coat color is defined as the dispersal of white hairs throughout the coat. In Quarter Horses, Paints, Welsh Ponies, and Belgians, roan has been associated with the *KIT* gene. Despite showing similar coloration, Clydesdales and Shire horses with roan coats do not have the roan allele observed in these other breeds. The purpose of this study was to identify a genomic region and underlying variation associated with roan in Clydesdale and Shire draft horses. For this study, 172 horses were sampled and phenotyped for the roan coat color. Phenotyping was performed by scoring the degree of roaning on the face, each leg, and the body on a four-point scale with zero being a solid

coat and three being extensive white hairs mixed into the coat. In addition to considering this as a quantitative trait, horses that scored either a zero or three (the extremes) for roaning on the body were used to model the phenotype as a binary trait. Genotyping was performed using the GGP Equine70K SNP Array. Genome-wide association studies (GWAS) were performed using Plink and GenABEL with some models considering other coat color genotypes (e.g., *MC1R*, *W20*) as covariates. Regions of interest identified by GWAS will be further investigated in additional samples not included in the initial data set and through whole genome sequence data of horses representing each phenotype.

PO0311: Equine

An 8.7kb Deletion in *MITF* Explains a Novel Splashed White Phenotype in the American Paint Horse

Jocelyn L. Tanaka, University of California - Davis, Davis, CA, Gary Magdesian, University of California Davis, Davis, CA and Rebecca R. Bellone, Veterinary Genetics Laboratory, University of California - Davis, Davis, CA

Splashed white is a coat color pattern in the horse characterized by extensive white patterning on the legs, belly, and face that is often accompanied by blue eyes and deafness. To date, three mutations, in *Microphthalmia-Associated Transcription Factor (MITF)* and two mutations in *Paired Box 3 (PAX3)* have been identified that explain splashed white patterns (SW1-5). An American Paint Horse stallion with a splashed white phenotype and blue eyes, whose parents were not white patterned, was unexplained by any of the five known splashed white variants or any other known white spotting patterns. We hypothesized this splashed white phenotype (SW6) was caused by a *de novo* mutation in *MITF* or *PAX3*. Analysis of whole genome sequencing using Illumina Novaseq technology with 150bp paired end reads to an average depth of 52X coverage identified an 8.7kb deletion in *MITF*. This variant removes 625 coding nucleotides, and is therefore predicted to impair protein function. No SNPs or structural variants were identified in *PAX3*. Sanger sequencing confirmed the stallion was heterozygous for the *MITF* deletion. Genotyping three of his splashed white offspring found that they were also heterozygous for the deletion. One additional offspring did not have the deletion and phenotypically and genotypically was identified as a frame overo like her dam. Given the role of *MITF* in producing white patterning phenotypes, and the predicted deleterious effect of this mutation, this 8.7kb deletion is the likely causal variant for SW6.

PE0312: Equine

Monoallelic Gene Expression; A Closer Look at Reciprocal Paternal and Maternal Gene Interaction in Equine Placenta

Pouya Dini¹, Peter Daels², Mariano Carossino³, Hossam El-Sheikh Ali⁴, Shavahn C. Loux⁵, Alejandro Esteller-Vico⁵, Kirsten E. Scoggin⁴, Ernest Bailey⁵, Ted Kalbfleisch⁶ and Barry A. Ball⁴, (1)Department of Veterinary Science, University of Kentucky / Faculty of Veterinary Medicine, University of Gent, Lexington, KY, (2)Faculty of Veterinary Medicine, University of Gent, Merelbeke, Belgium, (3)Louisiana State University, Baton Rouge, LA, (4)Department of Veterinary Science, University of Kentucky, Lexington, KY, (5)University of Kentucky, Lexington, KY, (6)University of Louisville, Louisville, KY

Development and function of the placenta, as the feto-maternal interface, depend on the dynamic and efficient gene expression in this tissue. Most of the autosomal genes in the placenta show a bi-allelic expression pattern. However, there are groups of genes that exhibit monoallelic expression in which allele-specific transcription is dependent on the parental origin of the chromosomes. It is believed that monoallelic gene expression (MAE) is involved in the reciprocal interaction between maternal and paternal genes, coordinating the allocation of resources between the fetus and the mother. One of the main challenges for studying MAE in the placenta is the maternal cellular contamination in the fetal part of the placenta. The horse has an epitheliochorial placenta in which both the endometrial epithelium and the epithelium of the chorionic villi are juxtaposed with minimal extension into the uterine mucosa. Yet, there is no information available on the allelic gene expression of equine chorioallantois (CA). In the current study, we present a dataset of 2,432 MAE genes in equine CA along with a workflow for analyzing monoallelic gene expression. We predicted several reciprocal, paternal and maternal interactions in gene expression of equine CA, including ligand-receptor and sense-antisense interactions. We further evaluated the expression

pattern of MAE throughout gestation. This study provides fundamental information regarding MAE during equine pregnancy, a species with a negligible amount of maternal contamination in its placenta. This information will provide the basis for further understanding the role of MAE during gestation

PO0313: Equine

Transcriptome Analysis Reveals the Key Regulators and Molecular Mechanisms Triggering Myometrial Activation during Equine Placentitis

Hossam El-Sheikh Ali^{1,2}, Pouya Dini³, Yatta Linhares Boakari¹, Shavahn C. Loux¹, Kirsten E. Scoggin⁴, Ted Kalbfleisch⁵ and Barry A. Ball⁴, (1)University of Kentucky, Lexington, KY, (2)Theriogenology Department, University of Manoura, Egypt, Mansoura, Egypt, (3)Department of Veterinary Science, University of Kentucky / Faculty of Veterinary Medicine, University of Gent, Lexington, KY, (4)Department of Veterinary Science, University of Kentucky, Lexington, KY, (5)University of Louisville, Louisville, KY

The better understanding of the molecular mechanisms underlying myometrial activation during equine placentitis is crucial to elucidate the pathways associated with preterm labor and the development of efficient therapies targeting these pathways. Therefore, the current study was designed to characterize the equine myometrium transcriptome during placentitis (290 d GA, n=6) in comparison to normal pregnant mares (290 d GA, n=6). Transcriptome analysis identified 255 DEGs (226 upregulated and 29 downregulated) in the myometrium during placentitis, including nine genes (*MMP8*, *MMP13*, *S100A9*, *S100A8*, *APOBEC3Z1B*, *PLA2G2D*, *CXCL17*, *LOC100061663*, and *LOC106782650*) that are exclusively expressed in the inflamed myometrium. To gain further insight into the biology of the DEGs, pathway analysis was carried out and revealed that these genes are involved in relevant pathways such as inflammation signaling (inflammation mediated by chemokine and cytokine signaling pathway, interleukin signaling, toll-like receptor signaling, and T cell activation), plasminogen activating cascade, and apoptosis signaling pathway, among others. Gene ontology enrichment analysis identified several chemoattractant factors in the inflamed myometrium, such as *CCL2*, *CXCL1*, and *CXCL17*. Upstream regulators analysis revealed 12 upstream regulators identified as DEGs, including transcription regulators (*CEBPD*, *SPI1*, *IRF7*, and *STAT1*), transmembrane receptors (*TLR2*, and *TYROBP*), enzymes (*PTGS2* and *PRKCP*), growth factor (*NRG1*), and cytokines and chemokines (*CSF3*, *S100A8* and *S100A9*). These findings revealed the key regulators and mechanisms triggering myometrial activation during equine placentitis, which might lead to the development of efficient therapies by targeting the key molecules and pathways.

PE0314: Equine

Comparing Microbiotas of Foals and Their Mares' Milk in the First Two Weeks after Birth

Michael J. Mienaltowski, Ubaldo De La Torre and Elizabeth A. Maga, University of California Davis, Davis, CA

The mare-foal relationship is essential for the well-being and growth of a foal. Mare's milk provides a foal with nutrients, protective immunity, and microbes. Within the first two weeks of life, there is a risk for a foal to suffer from diarrhea, particularly "foal heat diarrhea" which happens at about the time of a mare's estrus cycle but is more likely due to transitions in the microbiota in the foal's gastrointestinal (GI) tract. We hypothesized that this GI flora transition could be caused by changes in lysozyme and microbial populations in the mare's milk. To test this hypothesis, fifteen mare-foal pairs were followed in the first 15 days post-foaling. Every other day milk was collected from mares and rectal swabs were collected from foals. Mare's milk lysozyme activity peaked for samples at Day 1 and maintained levels between 72.5% (Day 15) and 86.5% (Day 3) of Day 1 activity. Microbial DNA was collected from the milk and swabs; the V4 domain of 16S rRNA genes were PCR amplified and sequenced using Illumina MiSeq technology. Microbial populations were analyzed using a combination of the QIIME2™ next-generation microbiome bioinformatics platform and Linear discriminant analysis Effect Size (LEfSe). Expected differences in microbial populations were found between mare's milk and rectal swabs -- *f.Enterobacteriaceae* and *f.Pasteurellaceae* vs. *f.Lachnospiraceae*, *f.Bacteroidaceae*, *f.Porphyrimonadaceae*, respectively. Further microbial population analyses are underway to better discern the relationship between microbes identified in mare's milk and those in a foal's GI tract over those first two weeks of a foal's life.

PO0315: Equine

Effect of Oral Urea Administration on the Transcriptome of the Equine Embryo

Yatta L. Boakari^{1,2}, Hossam El-Sheikh Ali^{1,3}, Pouya Dini^{1,4}, Shavahn C. Loux⁵, Claudia Barbosa Fernandes⁶, Kirsten E. Scoggin¹, Alejandro Esteller-Vico⁷, Laurie M. Lawrence⁸ and Barry A. Ball¹, (1)Department of Veterinary Science, University of Kentucky, Lexington, KY, (2)Clinical Sciences, Auburn University, Auburn, AL, (3)University of Mansoura, Mansoura, Egypt, (4)Department of Reproduction, Faculty of Veterinary Medicine, University of Gent, Merelbeke, Belgium, (5)University of Kentucky, Lexington, KY, (6)Department of Animal Reproduction, University of São Paulo, São Paulo, Brazil, (7)University of Tennessee, Knoxville, TN, (8)Department of Animal Science, University of Kentucky, Lexington, KY

High blood urea nitrogen (BUN) in ruminants has a negative impact on embryonic development; but, this relationship has not been reported in mares. However, high BUN concentrations have been shown to alter the endometrial transcriptome of mares. This study evaluated the effects of oral administration of urea to the mare on the transcriptome of day-14 equine embryos with RNA sequencing. When a 25±3 mm follicle was detected, mares were randomly allocated to a urea (n=9) or control treatments (n=10). The urea treatment consisted of an oral supplementation of urea (0.4 g/kg of body weight), mixed with sweet feed and molasses. The control treatment was sweet feed and molasses alone. Blood samples were collected every other day for BUN analysis. Mares were artificially inseminated in the presence of a 35-mm follicle and ovulation was detected (D0). Embryo collection was done at D14 (n=5 urea-treated; n=7 control embryos). Total RNA was extracted from embryos using TRIzol Reagent (RIN=8.95) for RNA sequencing and sequenced (2x150 base pairs/paired-end) with a NovaSeq 6000. The software STAR (2.5.3a) was used to map the reads using the EquCab3.0, then the ENSEMBL annotation was used with Cufflinks. Cuffdiff (2.2.1) was used to calculate differentially expressed genes (DEGs) between urea and control groups (FDR-adjusted p-value < 0.1). PANTHER (13.1) was used for functional annotation of the DEGs. There was an increase in BUN in the urea-treated group (P < 0.05). Additionally, a total of 14 DEGs were identified in embryos from urea-treated mares, such as neurofascin (*NFASC*), apelin receptor gene (*APLNR*), proprotein convertase (*PCSK1*) and prostate stem cell antigen (*PSCA*). The DEGs are involved in neurological development, cell proliferation, vascular remodeling, embryo adhesion and detoxification in the D14 embryos. In summary, oral urea treatment in mares caused transcriptomic changes on D14 equine embryos that might have deleterious effects to their development.

PE0318: Cattle

Genome Assemblies of Global Cattle Breeds to Create a Cattle Pangenome

Timothy P.L. Smith, USDA, ARS, USMARC, Clay Center, NE, Derek M. Bickhart, Dairy Forage Research Center, USDA-ARS, Madison, WI, Benjamin D. Rosen, Animal Genomics and Improvement Laboratory, USDA-ARS, Beltsville, MD and The Bovine PanGenome Project

The “gold standard” for conducting genomic research in any organism has been to first create a usable draft reference genome. However, it is increasingly recognized that a single reference genome, while very useful, is inadequate to fully describe the extent of genetic variation, even in relatively low-diversity species such as humans. This observation is the basis of the recently-announced Human Pangenome Reference Sequence Project (HPRSP), which aims to generate high-quality assemblies of the diversity of human genomes. Fortunately, advances in sequencing technologies have reduced the cost of creating assemblies to the point where even livestock species could reasonably achieve a similar goal if global resources can be marshaled to this end. The other big challenge to creating a useful Pangenome resource from these assemblies, is a method for representing genomes in such a way as to serve as a comprehensive reference map of genetic variation that includes all DNA segments existing in all members of the species. This latter challenge is a major initiative within HPRSP, and the tools developed there should be applicable to other species. We propose an initiative within the global cattle genomic research community, to generate reference-quality genomes for as many cattle breeds as possible. We suggest that the best strategy will make use of F1 crosses of breeds using the trio binning approach, and will outline a proposed framework to achieve a cattle pangenome in a time frame to coincide with development of visualization and analysis tools in the human genome community.

PO0319: Cattle**Mapping Sequencing Reads to Bovine Breed-Specific Genome Graphs****Danang Crysnanto** and Hubert Pausch, ETH Zurich, Zurich, Switzerland

The current bovine reference sequence is a linear consensus sequence derived from paternal and maternal haplotypes of a single highly-inbred Hereford cow. The linear sequence lacks diversity because it does not include allelic variation. Lack of diversity causes reference allele bias, i.e., DNA fragments that contain reference alleles are more likely to align correctly than those containing non-reference alleles. Variation-aware genome graphs may address problems arising from the linearity of current references. We used the existing bovine reference coordinate system (ARS-UCD1.2) as backbone and added sites of variation that were filtered according to dairy (Brown Swiss, Holstein) and dual-purpose (Fleckvieh, Original Braunvieh) cattle breeds to construct breed-specific genome graphs. Mapping accuracy did not differ between the linear reference sequence and genome graphs that were augmented with random variants. However, our results show that read mapping is more accurate to graph-based than linear reference genomes when the graph contains pre-selected variants. Variant prioritization is crucial to achieve high mapping accuracy at tractable computational complexity. Adding common variants improves mapping accuracy; but adding rare variants tends to compromise read mapping accuracy. Read mapping accuracy was the highest for breed-specific graphs i.e., when sequencing reads from Brown Swiss cattle were mapped to a variation-aware graph that was augmented with variants filtered according to the Brown Swiss population. We estimate that the number of erroneously mapped reads can be reduced by 1.5 million for a 35-fold coverage of the cattle genome when a variation-aware reference graph is considered. We anticipate that breed-specific genome graphs that were constructed from highly accurate and continuous breed-specific haplotype-resolved genome assemblies might further reduce mapping errors.

PE0320: Cattle**Annotation of Sequence Variants in the Bovine Genome with the Functional-and-Evolutionary Trait Heritability (FAETH) Score**

Ruidong Xiang¹, Irene van den Berg², Iona M MacLeod³, Ben J. Hayes⁴, Claire P. Prowse-Wilkins⁵, Min Wang⁶, Bolormaa Sunduimijid³, Zhiqian Liu², Simone Rochfort⁷, Coralie M. Reich³, Brett A. Mason³, Christy J Vander Jagt³, Hans D. Daetwyler³, Mogens Sandø Lund⁸, Amanda J. Chamberlain³ and Michael E. Goddard⁹, (1)University of Melbourne, Bundoora, VIC, Australia, (2)Agribio, Bundoora, VIC, Australia, (3)Agriculture Victoria, Bundoora, Australia, (4)University, Bundoora, VIC, Australia, (5)DJPR, Bundoora, Australia, (6)La Trobe University, Bundoora, Australia, (7)Agriculture Victoria, Bundoora, VIC, Australia, (8)Aarhus University, Tjele, Denmark, (9)The University of Melbourne, Parkville, Australia

The Functional Annotation of ANimal Genomes consortium provides comprehensive functional genomic information on farm animals. While many functional datasets are generated with different assays, it is unclear which types help identify causal genes. Moreover, it is unclear how to effectively analyse diverse functional datasets to locate causal mutations that will benefit genomic selection and precision animal breeding. We propose a framework which ranks millions of cattle sequence variants based on their 'Functional-And-Evolutionary Trait Heritability' (FAETH) score. We first collected 30 categories of functional and evolutionary data in cattle and other species, including metabolic quantitative trait loci (QTLs), splicing QTLs, ChIP-seq peaks, selection signatures and conserved sites across 100 vertebrate species. Then, we partitioned genetic variance of 34 cattle traits into these functional categories with genome-wide restricted maximum likelihood analysis in over 44,000 Australian dairy bulls and cows with 17.7 million imputed sequence variants. Based on the per-variant trait heritability and the variant memberships to different functional categories, we rank both the importance of different functional datasets and individual variants. Validated in over 7,700 Danish cattle, the high FAETH-ranking variants showed significantly increased genetic variances and genomic prediction accuracies. Our study provides 1) proof-of-concept evidence that functional genomics can improve genomic selection in global cattle breeds; 2) novel and effective methods for integrating functional data into genome-wide analysis of sequence variants and 3) the publicly available FAETH score of 17.7 million cattle variants which can be used as biological prior for genomic selection or new functional annotation resources.

PO0321: Cattle

Identification and Characterization of microRNA Genetic Variants in Dairy Cattle, From Their Detection to the Analysis of Their Biological Impacts

Céline Bourdon¹, Mekki Boussaha¹, Marie-Pierre Sanchez¹, Thierry Tribout¹, Rachel Lefebvre¹, Sandrine Le Guillou¹, Gwenola Tosser² and Fabienne Le Provost¹, (1)INRA UMR1313 Animal Genetics and Integrative Biology, Jouy en Josas, France, (2)INRA GenPhySE, Castanet-Tolosan cedex, France

Genomic selection is now widespread in bovine species, leading to the selection of animals through their Estimated Breeding Value (EBV), taking into account relevant traits. The addition of causal mutations to anonymous genetic markers could improve EBV accuracy. To participate in this improvement in a dairy context, genetic variants of microRNAs expressed in mammary gland and localized in dairy QTL were studied in bovine.

Starting from millions of genetic variants from whole genome sequencing data, we selected those i) in a genomic region significant for dairy traits and ii) in a microRNA expressed in mammary gland. Three of them were validated thanks to GWAS data, with a validated link between genotype and phenotype. Biological impacts of the validated variants were analyzed according to their expected effect. The expression level of the microRNA was studied if its biogenesis was thought to be impacted, and the expression levels of targeted mRNAs was studied if the impact was expected on the microRNA/mRNA recognition. Notably, modifications of targeted mRNAs expression levels were observed, emphasizing the impact of a single nucleotide change in the mRNAs recognition.

These steps lead to an integrated pipeline for the analysis of microRNA genetic variants. Thanks to its validation through the achieved results, the developed approach will be applied to ovine and caprine datasets.

PE0322: Cattle

Parent-Offspring Genotyped Trios Unravelling Regions with Allelic and Genotypic Epistatic Transmission Ratio Distortion on the Cattle Genome

Samir Id Lahoucine¹, Joaquim Casellas², Pablo A. S. Fonseca³, Aroa Suarez-Vega¹, Flavio Schenkel³, Stephen Miller⁴, Mehdi Sargolzaei⁵, Michael Lohuis⁶, Juan F Medrano⁷ and Anegla Canovas⁸, (1)Centre for Genetic Improvement of Livestock, Department of Animal Biosciences, University of Guelph, Guelph, ON, Canada, (2)Universitat Autònoma de Barcelona, Departament de Ciència Animal i dels Aliments, Barcelona, Spain, (3)University of Guelph, Guelph, ON, Canada, (4)Angus Genetics Inc., St. Joseph, MO, (5)Select Sires, Inc., Plain City, OH, (6)The Semex Alliance, Guelph, ON, Canada, (7)University of California-Davis, Davis, CA, (8)Department of Animal Biosciences, University of Guelph, Guelph, ON, Canada

Several biological mechanisms affecting the sperm and the ova fertility and the viability at developmental stages of the reproductive cycle resulted in observable deviation from Mendelian expectations (i.e., transmission ratio distortion, TRD). Although the genetic background of TRD has been mainly linked to single-locus factors, gene-by-gene interactions (i.e., epistasis) could also potentially cause specific TRD patterns. Thus, for example, compatible allele (or genotype) combinations at different loci increase fitness whereas unfavorable allelic combinations are under-represented, exhibiting signals of TRD. It has also been hypothesized that the incomplete penetrance of lethal mutations may involve epistatic interaction. Here, we aimed to uncover and explain the incomplete deviation of single-locus TRD using an epistatic TRD model and identify new pairs of genomic regions with epistatic TRD in Holstein cattle. Allelic and genotypic parameterizations for direct and epistatic TRD were developed and implemented to scan the whole genome in inter-chromosomal SNP pairs using 283,817 sire-dam-offspring genotyped trios with 47,910 SNPs. The allelic TRD model identified 17 pairs of epistatic SNPs, explaining up to 0.25 (50%) additional TRD rate compared to single-locus TRD, which may be associated with the low penetrance of single-locus distortions. Using the genotypic model, after the multiple test correction, 7, 19 and 6 regions were found with decisive evidence ($\log_{10}(\text{Bayes Factor}) > 1000$) with additive-by-additive, additive-by-dominance and dominance-by-dominance effects, respectively. Therefore, this study showed that different epistatic TRD patterns exist in the Holstein genome and presented a list of new candidate genomic regions that show epistatic interactions with potentially important biological implications.

PO0323: Cattle

Bootstrapping Population in Cattle GWAS

Jenna A. Kalleberg¹, Troy N. Rowan², Sara M Nilson² and Robert D. Schnabel³, (1)Institute for Data Science and Informatics, Division of Animal Science, University of Missouri, Columbia, MO, (2)Division of Animal Sciences, University of Missouri, Columbia, MO, (3)Division of Animal Sciences, Informatics Institute, Columbia, MO

Substantially removing noise in Genome-Wide Association Studies (GWAS) would improve our confidence in detecting Quantitative Trait Loci. However, our inability to recognize genuine associations from false-positives using empirical evidence remains a barrier to accurate interpretation. We evaluated how using a different set of individuals impacts the persistence of marker effects. Through bootstrapping sample size, our approach provides additional evidence to classify the Type I error rate of associations discovered from GWAS. Using a cattle feed efficiency dataset with 11,697 individuals with imputed genotypes and multiple phenotypes, we created 100 sample-size permutations, each with 5,000 individuals. Using all individuals, we identified *potential positive markers* by applying the Fixed and random Circulating Probability Unit (FarmCPU) model. We identified *repeating positive markers* using the same model for our permuted samples. To improve the confidence in determining true associations, we compared the results from all individuals against the outcomes from the modified sample size analyses. *Potential positive markers* that repeatedly occurred in the sub-sample analyses were labeled true positives. By calculating the frequency of marker recurrence across the 100 sub-samples, we reassure our classification of markers as true or false positives. Next, we identified the ratio of our *potential positives* that may be false. In summary, we present how varying sample size provides evidence for evaluating markers. Our workflow allows iterative GWAS analysis using randomly generated sub-samples of individuals. By bootstrapping sample size, we can improve the interpretation of GWAS analyses.

PE0324: Cattle

Genome Wide Analysis of Antral Follicle Counts in Crossbred Heifer

Morgan R. Stegemiller¹, Kimberly M. Davenport¹, Maggie K. Reynolds¹, Troy N. Rowan², John B. Hall¹ and Brenda M. Murdoch¹, (1)Department of Animal and Veterinary Science, University of Idaho, Moscow, ID, (2)Division of Animal Sciences, University of Missouri, Columbia, MO

The ability of heifers to conceive earlier contributes to the efficiency and profitability of cattle breeding operations. Heifers with a higher antral follicle count (AFC) are more likely to conceive and tend to have a longer reproductive span and produce more offspring. Determining genetic associations with AFC can help producers select more reproductively efficient heifers. Blood samples were collected, and DNA extracted for 216 crossbreed beef heifers, 70 from year one and 146 from year two. Genotypes were obtained using a 50K Single Nucleotide Polymorphism (SNP) marker array and then imputed up to 850,000 SNP markers. A principal component analysis plot, used to correct for relatedness, showed four separate groupings which corresponded with the sire's breeds. Using a single-locus mixed model with year, age at time of sampling, and the four PCA groups as covariates an association analyses was performed. Two significant SNPs were identified on chromosome 23 ($P \leq 8.33e-8$) and one on chromosome 2 ($P \leq 2.89e-7$). An increase in AFC in heifers can increase the rate of conception as well as the number of offspring produced over a lifetime. Determining a genetic association with AFC will enable producers to select for animals with increased reproductive performance and have a more efficient production process.

PO0325: Cattle

Rumen Epithelial Transcriptome and Microbiome Profiles of Beef Cattle with Liver Abscesses.

Amanda K. Lindholm-Perry¹, Waseem Abbas², Brittney N. Keel¹, Samodha C. Fernando³, James E. Wells⁴ and Kristin E. Hales⁵, (1)USDA, ARS, U.S. MEAT ANIMAL RESEARCH CENTER, CLAY CENTER, NE, (2)University of Nebraska Lincoln, Lincoln, NE, (3)University of Nebraska, Lincoln, NE, (4)USDA-ARS, U.S. Meat Animal Research Center, Clay Center, NE, (5)Texas Tech University, Lubbock, TX

Liver condemnation is most often caused by liver abscess and is estimated to cost the beef industry \$64 million annually. *Fusobacterium necrophorum*, a bacterial species commonly found in the rumen, is the primary pathogen associated with liver abscess in cattle by transport to the liver via damage to the rumen wall. The purpose of this study was to determine whether there were gene expression and microbial population differences in the rumen epithelium of beef cattle with severe liver abscesses compared to those without liver abscess. Rumen tissue was collected from 31 beef steers and heifers with liver abscess and 30 animals with no liver abscess for the evaluation of gene expression differences. A total of 221 genes were identified as differentially expressed in the animals with severe liver abscess. Genes belonging to the nuclear factor kappa-light-chain-enhancer of activated B cells (NF- κ B) signaling and interferon signaling pathways were identified as over-represented in the list of differentially expressed genes (DEG). Most of these genes were down-regulated in animals with severe liver abscesses. There were differences in the microbial communities of the rumen papillae between animals with and without liver abscesses. In addition, strong correlations were detected between specific epithelial bacterial groups and DEG. These data suggest that there are both host and microbial factors impacting gene expression in the rumen papillae, which may affect the development or persistence of liver abscesses in beef cattle.

PE0326: Cattle

Association of *ARRDC3* and *NF1A* Genes with Bovine Congestive Heart Failure

Michael P. Heaton¹, Gregory P. Harhay¹, Adam S. Bassett², Halden J. Clark², Jaden M. Carlson², Helen R. Smith¹, Madeline C. Pelster², Aspen M. Workman¹, Dale M. Grotelueschen² and Brian L. Vander Ley², (1)USDA, ARS, U.S. Meat Animal Research Center (USMARC), Clay Center, NE, (2)University of Nebraska, Lincoln, Clay Center, NE

Background: Bovine congestive heart failure (BCHF) has become increasingly prevalent in feedlot cattle in the Western Great Plains of North America with up to 7% mortality in disease outbreaks. BCHF is an untreatable complex condition involving pulmonary hypertension that culminates in right ventricular failure and death. No candidate genes are presently associated with BCHF, and thus, our aim was to search genome-wide for genetic risk factors in feedlot cattle.

Methods: Samples of 102 clinical BCHF cases and 102 unaffected matched penmates were used in a genome-wide association study (GWAS) with 777,962 single nucleotide polymorphisms (SNPs). The paired nominal data were analyzed with McNemar's test.

Results: Analyses of 563,042 filtered SNPs revealed more than 15 genomic regions highly associated with BCHF. Regions with the strongest association included the arresting domain-containing 3 protein (*ARRDC3*) and nuclear factor 1A (*NF1A*) genes. A missense mutation in exon 4 of *ARRDC3* (C182Y), and SNPs in intron 5 of *NF1A* had the best statistical support for association (McNemar's Chi-square > 20). Animals with either or both the *ARRDC3* or *NF1A* risk factors, were approximately 7- and 15-fold more likely to have BCHF compared to those without (p -value < 10^{-10} for both risks present). A two-SNP genotyping test for *ARRDC3* and *NF1A* risk factors was used to test an independent cohort of feedlot cattle with end-stage heart failure and similar associations with disease were observed.

Conclusions: A matched case-control GWAS identified major genes associated with BCHF in feedlot cattle. Although the roles of these genes in disease pathogenesis are unknown, their discovery facilitates classifying animals by genetic risk for heart failure and will allow producers to make informed decisions for selective breeding and animal health management.

PO0327: Cattle

Effect of Toxic Fescue on Whole Blood Gene Expression in Beef Cattle

Marcela M deSouza¹, Hamid Beiki¹, Dawn Koltes¹, Jeremy G Powell², Julie Atchley², Laura Meyer², John Tucker^{2,3}, Donald S Hubbell, III^{2,3} and James E. Koltes¹, (1)Department of Animal Science, Iowa State University, Ames, IA, (2)Department of Animal Science, University of Arkansas, Fayetteville, AR, (3)Livestock & Forestry Research Station, Division of Agriculture, University of Arkansas, Batesville, AR

Endophyte-infected toxic fescue impacts 20% of the cattle in the United States due to its heat/drought tolerance and high nutrient qualities. However, the endophyte produces mycotoxins which cause increased heat stress in cattle as well as negative effects on growth, reproduction, and health, known as fescue toxicosis. Understanding how gene expression is impacted by toxic fescue may improve our diagnosis of fescue toxicosis and provide clues about the genetic mechanisms involved in cattle. The goal of this study was to identify differentially expressed genes (DEG) in the blood of cows grazed on toxic, non-toxic, or alternating pastures to identify biomarkers of fescue toxicosis. Blood samples were collected for RNA-sequencing (Illumina HiSeq3000, 150 nt, single-end). Sequence quality was evaluated with FastQC and adaptors/reads were trimmed using BBDuk. Sequences were aligned to the ARS-UCD1.2 genome using STAR software and quantified with Featurecounts. Reads were filtered for non/lowly genes, resulting in 80 samples for analysis. Statistical analysis was conducted using PROC Glimmix (SAS), where TMM normalized expression were adjusted for pasture type, breed, parity, pregnancy, lane and overdispersion variables. Q-values were used to control the false discovery rate ($q < 0.05$), resulting in 605 DEGs. Functional analysis (ClueGO) identified animals from non-toxic pasture were enriched for KEGG pathways as protein processing and antigen processing and presentation. Genes highly expressed on toxic pasture were over-represented for a wide range of signaling pathways as prolactin, oxytocin, GnRH, apelin, glucagon, B cell receptor, and others pathways as apoptosis, platelet activation and growth hormone secretion.

PE0328: Cattle

DNA Methylation in the Regions of Structural Variation in the Limbic System of Cattle

Suraj Bhattarai¹, Bonnie Cantrell¹, Brenda Murdoch², Rick Funston³, Robert Weaber⁴ and Stephanie McKay¹, (1)University of Vermont, Burlington, VT, (2)Department of Animal and Veterinary Science, University of Idaho, Moscow, ID, (3)University of Nebraska, North Platte, NE, (4)Kansas State University, Manhattan, KS

Genomic alterations larger than 50bp comprise structural variations (SVs), which include insertions, deletions, inversions, and translocations in the genome. SVs can explain greater proportion of phenotypic variation as they cover greater than 10% of the genome compared to 0.1% by single nucleotide polymorphisms (SNPs) in mammals. In cattle, DNA methylation and SVs, separately have been shown to be associated with various phenotypic traits. However, the effect of both DNA methylation and SVs in phenotypic traits have not been examined. Understanding the profile of DNA methylation in the SVs will uncover the extent to which SVs influence genomic function and how SVs arise and are being maintained in the genome. The objective of this study is to determine the average methylation level in varying annotated genomic regions (e.g. promoters, genic regions, intergenic regions and CDS); average methylation levels throughout the whole genome, and average methylation levels only in the regions which are captured by discordant mapping; which corresponds to the structural variations. To this end, we have whole genome bisulfite sequenced five regions of the limbic system of the bovine brain in each of 8 Red Angus x Simmental steers with extreme measures of docility. Sequence reads were mapped to the bovine reference genome ARS-UCD1.2 using the BSSeeker2 pipeline and discordant reads, which represents the regions of structural variation, will be used to generate methylation profiles. We aim to find a significant association between DNA methylation in SV regions which might explain some of the variation in cattle with extreme measures of docility.

PO0329: Cattle

Identification of 5-Hydroxymethylcytosine Markers in the Cattle Brain

Emily Stassen¹, Brenda Murdoch², Rick Funston³, Robert Weaber⁴ and Stephanie McKay¹, (1)University of Vermont, Burlington, VT, (2)Department of Animal and Veterinary Science, University of Idaho, Moscow, ID, (3)University of Nebraska, North Platte, NE, (4)Kansas State University, Manhattan, KS

Epigenetic variation plays an important role in phenotypic variability. DNA methylation is a dynamic epigenetic marker that can be added and removed from DNA nucleotides. The removal pathway begins when 5-methylcytosine (5-mC) is oxidized to 5-hydroxymethylcytosine (5-hmC). Both 5-mC and 5-hmC are stable epigenetic markers that influence transcription. 5-mC is thought to inhibit transcription, whereas 5-hmC is thought to promote transcription. It is important to distinguish between these stable epigenetic markers to gain a clear understanding of how epigenetics is influencing transcription. However, traditional whole genome bisulfite sequencing methods do not differentiate between these markers. Because of this, 5-hmC is often misrepresented as 5-mC. The grouping of these

two distinct markers during data collection could lead to misinformation on the impact of DNA methylation markers on transcription and phenotype. Differentiating between markers is especially significant in the brain, where there is an increase in the presence of 5-hmC. We have previously studied the relationship between DNA methylation and docility using standard whole genome bisulfite sequencing on select brain tissues. To differentiate between 5-hmC and 5-mC, we have carried out reduced representation oxidative bisulfite sequencing on brain tissues from the limbic system of 8 Red Angus x Simmental steers. Fastq sequencing files were trimmed with Trim Galore. Subsequently, Bismark was used for alignment to bovine reference index ARS-UCD 1.2 and to call hydroxymethylation. These data show the presence of 5-hmC in the bovine brain, which emphasizes the need to differentiate between 5-hmC and 5-mC.

PE0330: Cattle

Assessing the Potential of Low-Pass Sequencing and Imputation to Genotype Sequence Variants in Beef Cattle

Warren M. Snelling¹, Larry A. Kuehn¹, Jeremiah H. Li², Jesse L. Hoff² and Joseph K. Pickrell²,
(1)USDA, ARS, U.S. MEAT ANIMAL RESEARCH CENTER, CLAY CENTER, NE, Clay Center, NE,
(2)Gencove Inc, New York, NY

Low-coverage whole-genome sequence (WGS) can be obtained at a cost similar to the low cost, low density cattle genotyping arrays. While few genotypes can be called directly from the <1X coverage that might be acquired, sequence variant genotypes have been imputed with high accuracy from <1X coverage of human samples. For an initial assessment of low-coverage sequencing and imputation in beef cattle, reads were sampled from bulls with >4X WGS to mimic low-pass sequencing. Each bull was downsampled to five coverage levels: 0.4X, 0.6X, 0.8X, 1X and 2X. The downsampled fastq files were submitted to the Gencove pipeline for imputation to a broad reference of *Bos taurus*, *Bos indicus* and *indicus*-influenced composites. Imputed genotypes were compared to genotypes from the BovineHD (HD) and GGP-F250 (F250) assays, as well as to genotypes called directly from deeper sequence on these bulls. Sequence reads were from eleven bulls, one representing each of the seven most predominant *Bos taurus* breeds in the U.S., a Brahman, and one representing each of three *indicus*-influenced composite breeds. Genotypes were imputed for 59M variants with 588K expected to alter coded proteins. For all bulls, agreement between imputed and assayed genotypes increased slightly with coverage. *Bos taurus* bulls consistently had concordance >0.99 between imputed and assayed genotypes. Concordance for the *indicus* x *taurus* composite bulls was between 0.96 and 0.97, and the Brahman was intermediate between the *taurus* and composite bulls. These results encourage use of low-pass sequencing and imputation as a viable alternative to genotyping assays.

PO0331: Cattle

Accuracy of Indirect Predictions Based on Prediction Error Covariance of SNP Effects From Single-Step GBLUP

Andre Garcia¹, Ignacio Aguilar², Andres Legarra³, Stephen Miller⁴, Shogo Tsuruta¹, Ignacy Misztal¹ and Daniela Lourenco¹, (1)University of Georgia, Athens, GA, (2)INIA, Montevideo, Uruguay, (3)INRA, Toulouse, France, (4)Angus Genetics Inc., St. Joseph, MO

In large genotyped populations, the value of adding all genotyped animals into the single-step GBLUP (ssGBLUP) evaluation is unclear. Indirect predictions (IP) are useful to compute genomic predictions for genotyped animals not included in the evaluations. IP can be calculated as the sum of SNP effects weighted by the gene content. Having IP is beneficial if their accuracy is similar to GEBV accuracy. The objective of the study was to compute accuracy for IP using the direct inversion of the genomic relationship matrix or the algorithm for proven and young (APY). Accuracy for IP was calculated by back-solving prediction error covariance (PEC) of GEBV into PEC of SNP effects. Data were provided by the American Angus Association and consisted of 35k post-weaning gain phenotypes and pedigree information for 192k animals. Out of these, 60k were genotyped for 38k SNP after quality control. A complete dataset with phenotypes and pedigree up to 2013 and genotypes up to 2014 was used to obtain GEBV accuracy as benchmark. A reduced dataset had the same phenotypes and pedigree, but genotypes of 2k validation animals born in 2014 were omitted. IP and their accuracies were calculated for validation animals based on reduced dataset. Correlations between benchmark and IP accuracy were greater than 0.99, with or without APY. The results

indicate that a good measure of accuracy for IP can be obtained when using ssGBLUP with and without APY and ongoing studies aim to extend it to larger genotyped populations.

PE0332: Cattle

Blood Traits in Beef Cattle and Their Relationship with Production Traits at Weaning

Josue Chinchilla Vargas¹, Luke Kramer¹, John Tucker², Donald S Hubbell, III³, Jeremy G Powell⁴, Toby D Lester⁴, Elizabeth A Backes⁴, Karen Anschutz⁴, Jared E. Decker⁵, Kenneth J Stalder¹, Max F. Rothschild¹ and James E. Koltes¹, (1)Department of Animal Science, Iowa State University, Ames, IA, (2)Livestock & Forestry Research Station, Division of Agriculture, University of Arkansas, Batesville, AR, (3)Division of Agriculture, University of Arkansas, Batesville, AR, (4)Department of Animal Science, University of Arkansas, Fayetteville, AR, (5)Division of Animal Sciences, University of Missouri, Columbia, MO

Disease represents one of the main factors that determine profitability in animal production. Previous research has observed significant correlations between blood cell counts and the animal's health status. We hypothesize that blood cell traits may be an effective indicator of performance in beef cattle. Complete blood counts were recorded from approximately 500 crossbred animals at weaning (Angus background crossed with Hereford, Charolais, Sim-Angus, Brangus) born between 2015 and 2016 and raised on toxic or novel tall fescue on three different farms. The animals were genotyped at an approximate density of 50,000 SNPs and the genotypes were imputed to an approximate density of 200,000 SNPs. Heritability, genetic and phenotypic correlations were estimated for 15 blood and 4 production traits. Additionally, with the objective of identifying the genetic basis underlying the different blood traits, a genome-wide association study (GWAS) was performed for all traits. Heritability estimates ranged from 0.11 to 0.60, and generally weak phenotypic correlations and strong genetic correlations were found. GWAS identified 90 1 Mb windows that explained 0.5% or more of the estimated genetic variance for at least 1 trait with 21 windows overlapping two or more traits. Further research efforts include identifying underlying candidate genes for traits and comparing toxic and novel fescue effects on blood traits. It appears that blood traits have weak phenotypic correlations but strong genetic correlations among themselves, as evidenced by important overlapping regions of genetic control for similar blood traits. However, blood traits have limited potential as indicator traits for productivity.

PO0333: Cattle

Regulation of Metallothionein Family Gene Expression in Nellore Cattle Evaluated for Residual Feed Intake

Larissa F. S. Fonseca¹, Danielly B. S. Silva¹, Patrícia I. Schmidt¹, Maria E. Z. Mercadante^{2,3}, Sarah F. M. Bonilha^{2,3} and Lucia G. Albuquerque^{1,3}, (1)São Paulo State University, School of Agricultural and Veterinarian Sciences – FCAV/UNESP, Jaboticabal, SP, Brazil, (2)Beef Cattle Research Center, Animal Science Institute, Sertãozinho, Brazil, (3)National Council for Scientific and Technological Development, Brasília, Brazil

The RNA sequencing (RNA-Seq) was used to identify differentially expressed genes in Nellore cattle liver tissue evaluated for residual feed intake (RFI). The analysis included liver samples from 24 animals significantly divergent (t-test<0.05) for RFI, being 12 more efficient (-0.45±0.18 kg DM/d) and 12 less efficient (0.39±0.19 kg DM/d). GeneMania and CentiScaPe apps, both implemented in Cytoscape software, were used to generate an interaction network from 53 differentially expressed genes, obtained in the RNA-Seq analysis, and to characterize the overall network topology using Node Degree and EigenVector parameters, respectively. Three metallothioneins family genes, MT1A, MT1E and MT2, were found as highly connected nodes. They were more expressed in the low efficient animals. This family includes cysteine-rich heavy metal-binding proteins and can act in protecting oxidative stress-inducing agents and toxic metals. Transcription of these genes is rapidly up-regulated in response to agents that cause oxidative stress. The RFI has been associated with mitochondrial function since phosphorylation of ADP, main energy source of the cell, occurs in these organelles. A depletion in ATP synthesis results from the increasing of generation of reactive oxygen species, due oxidative stress. Therefore, the highest expression of these genes in the low efficient animals indicates a cell response to the oxidative stress-inducing agents, suggesting an important mechanism that induces the biochemistry redress of the lost feed efficiency. Acknowledgments: São Paulo

Research Foundation (Grants #2009/16118-5, #2017/10630-2, and #2019/16732-7) and Coordination for the Improvement of Higher Education Personnel - Brasil (CAPES - #001 and #88887.337783/2019-00).

PE0334: Cattle

Identification of Candidate Lethal Haplotypes in Nellore Cattle

Patrícia I. Schmidt¹, Paul VanRaden², Li Ma³, Daniel J. Null², John B. Cole², Gabriel S Campos⁴, Daniel J. A. Santos³, Ana F B Magalhães¹, Larissa F. S. Fonseca¹, Roberto Carneiro¹, Arione A. Boligon⁵, Gerardo A. Fernandes Júnior¹ and Lucia G. Albuquerque¹, (1)São Paulo State University, School of Agricultural and Veterinarian Sciences – FCAV/UNESP, Jaboticabal, SP, Brazil, (2)Animal Genomics and Improvement Laboratory, USDA-ARS, Beltsville, MD, (3)Department of Animal and Avian Sciences, University of Maryland, College Park, MD, (4)Department of Animal and Dairy Science, University of Georgia, Athens, GA, (5)Federal University of Pelotas, Animal Science Department, UFPel, Capão do Leão, RS, Brazil

With genomic data, lethal recessives may be discovered from haplotypes that are common in the population but are never homozygous in live animals. The objective of the present study is to identify lethal haplotypes in a Nellore population, based on expected population frequencies. The method requires genotypes only from apparently normal individuals and not from affected embryos. A total of 4,447 animals were genotyped with a high-density panel (777,962 SNP markers) and 4,041 with a panel containing 74,677 markers, which were imputed to the HD panel using Findhap software v3. Map locations are from the ARS-UCD1.2 *Bos taurus* genome assembly. The program first examined haplotypes of 2,000 markers, then 632 markers, and finally identified haplotypes of ≤ 200 markers for further analysis. Expected numbers of homozygous individuals were calculated through two methods: Simple - assumed random mating and used the number of individuals genotyped divided by 4 and multiplied by the square of the carrier frequency; Mating - used the actual mating pattern and was the number of carrier service sire \times carrier maternal grandsire matings divided by 4. Thirty haplotypes had high frequency expected but no homozygotes observed. Of these, the haplotypes with most expected homozygotes were on chromosomes 23 at location 26,021-1,523,616 (simple method - 38.6; mating method - 22.5) and 19 at location 58,555,500-59,418,915 (simple method - 27.0; mating method - 21.5). These tests to identify lethal haplotypes can help breeders to anticipate problems caused by disadvantageous variants through the implementation of selection actions and management of matings. Acknowledgments: São Paulo Research Foundation (Grants #2009/16118-5, #2017/10630-2, #2018/17812-1 and #2019/10123-9) and Coordination for the Improvement of Higher Education Personnel - Brasil (CAPES - #001).

PO0335: Cattle

Nutrigenetic Effect of Fetal Programming on Beef Cattle Performance at Weaning

Miguel H. A. Santana, Fernando O. Bussiman, Guilherme H. G. Polizel, Roberta C. Cracco, Edison Furlan, Isabela Mortari, Felipe E. Carvalho, Bárbara C. Abreu Silva and Jose B. S. Ferraz, University of Sao Paulo/College of Animal Science and Food Engineer/FZEA, Pirassununga, Brazil

Indications that malnutrition at critical stages of prenatal and postnatal life have permanent effects on livestock health and productivity are known. However, there is still much to study about the genetic effects of these nutritional stimuli/deficiencies during pregnancy, especially the quantitative importance of the phenomenon in relation to their effects on the productive efficiency in the adult life of cattle. The aim of this study was to evaluate the nutrigenetic effect of fetal programming on the performance of Nellore cattle. Weaning weight (WW) and average weight gain (AWG) of 120 calves were used from three treatments of nutrition from their mothers during pregnancy: NP - only mineral supplementation, PP - protein-energy supplement in the final third and PC - protein-energy supplement the entire pregnancy. Pedigree and genotypic data from these animals and their progenitors were used to evaluate genotype by environment (GE) interactions through genomic correlations between the genomic breeding values (GEBV) of the animals in the three treatments using sex as a fixed effect, cow and calf age as linear covariate. For WW the correlations were: NPxPP = 0.18, NPxPC = -0.17 and PPxPC = 0.05. For AWG were: NPxPP = 0.26, NPxPC = -0.39 and PPxPC = 0.09. Correlation of GEBV for a trait in different environments smaller than 0.8 is an indicate of GE, thus the found results indicate that there may be evidence of GE for fetal programming in cattle. The nutrigenetic effects seems to affect the performance of beef cattle.

PE0336: Cattle

Analysis of Divergent Transcription Factors across Different Cattle Tissues Reveals Their Interesting Biological Functions

Hamid Beiki¹, James E. Koltes², Zhiliang Hu¹ and James M. Reecy¹, (1)Department of Animal Science, Iowa State University, Ames, IA, (2)Iowa State University, Ames, IA

Divergent transcription, defined as two polymerases transcribing on opposite directions from the same promoter, is a wide-spread phenomenon in mammals. Many independent experiments indicate that transcription can occur divergently from many promoters in mammals. This phenomenon is significantly associated with genes related to transcriptional regulation and development. In this study, we used single-molecule long-read isoform sequencing (Iso-seq) and strand-specific Illumina RNA sequencing (RNA-seq) to identify antisense transcripts from gene promoters in cattle tissues. Our preliminary results are based on 101 high-quality strand-specific RNA-seq data from untreated, healthy Holstein cattle from four tissues (liver, macrophage, ovary and adipose). This analysis revealed that more than 80% of the sense/antisense gene pairs with significant expression correlation (q -value <0.05) were positively correlated in all tissues, except liver ($>60\%$ of significant correlations were negative). This result was consistent across all possible combination of sense/antisense gene pair biological types (coding-coding, coding-lncRNA and lncRNA-lncRNA). In addition, more than 50% of sense/antisense gene pairs with significant correlations in liver also had significant reverse correlations in macrophages, and the ratio was 32% for the same comparison between liver and ovary. These preliminary findings seem to indicate that (1) mechanisms may exist to cause these divergent expression differences, (2) the massive divergent transcription differences in cattle tissues might play important roles in terms of developmental biology.

PO0337: Cattle

lncRNA IGF2 AS Regulates Bovine Myogenesis Through Different Pathways

Chengchuang Song¹, ChunLei Zhang¹, Xingtang Fang¹ and Hong Chen^{1,2}, (1)Jiangsu Normal University, Xuzhou, China, (2)Northwest A&F University, Yangling, China

The role of long non-coding RNA (lncRNA) in the regulation of bovine skeletal muscle development remains poorly understood. The present study investigated the function and regulatory mechanism of a novel lncRNA, IGF2 AS, in bovine myoblasts proliferation, differentiation and apoptosis. Gain or loss of IGF2 AS was performed using an expression plasmid or siRNA, respectively. Bovine myoblasts were used to investigate the biological function and mechanisms of IGF2 AS in vitro. Results were conjoint analyzed by cellular and molecular biology experiments as well as bioinformatics. Functionally, IGF2 AS could promote proliferation and differentiation of bovine myoblasts and inhibit apoptosis. Preliminary mechanism suggests, on the one hand, IGF2 AS could complement the IGF2 gene intron region and affect the stability and expression of IGF2 mRNA. On the other hand, RNA pull-down and immunoprecipitation assays demonstrated that IGF2 AS could directly bind to the ILF3 protein and maybe partly though it to regulate myogenesis. In conclusion, the novel identified lncRNA IGF2 AS promoted proliferation and differentiation and inhibited apoptosis of bovine myoblasts through various pathways.

PE0338: Cattle

miR-204-5p Regulates Milk Synthesis in Mammary Epithelial Cells through ATF2

Molan Zhang, Jia Liu, YanHong Wang, Chengchuang Song, Xingtang Fang and **Chunlei Zhang**, Jiangsu Normal University, Xuzhou, China

The role of microRNA in the regulation of milk synthesis in mammary gland remains poorly understood. The present study investigated the function and regulatory mechanism of miR-204-5p in mammary epithelial cells. The mimics of miR-204-5p promoted lipid synthesis, while knock-down of miR-204-5p by inhibitor depressed lipid synthesis in vitro in mouse mammary epithelial cells (HC11). Through a combination of bioinformatics analysis, target gene 3' UTR luciferase reporter assays, and western blotting, we identified activating transcription factor 2 (ATF2) as a target of miR-204-5p. Transfection of siRNA-ATF2 into HC11 led to increases of triglyceride, suggesting a negative role of ATF2 in mammary epithelial cell lipid synthesis. In summary, data suggest that miR-204-5p regulates biological processes associated with intracellular triglyceride synthesis through ATF2. These data

provide a theoretical and experimental framework for further clarifying the regulation of lipid metabolism in mammary gland.

PO0339: Cattle

A Functional 3'UTR Polymorphism of FADS2 Affects Cow Milk Composition Through Modifying miR-744 Binding

Mingxun Li^{1,2}, Zhangping Yang², George E. Liu¹ and Yongjiang Mao², (1)Animal Genomics and Improvement Lab. BARC, NEA, ARS, USDA, Beltsville, MD, (2)College of Animal Science and Technology, Yangzhou University, Yangzhou, China

Fatty acid desaturase 2 (FADS2) is the rate-limiting enzyme involved in the synthesis of long-chain polyunsaturated fatty acids (LC-PUFAs). Many studies have suggested that polymorphisms in the *FADS2* gene can modify its delta-6 desaturase activity. However, much remains unknown in regards to the regulatory mechanisms interpreting how DNA variants influence the function of FADS2. In the present study, we identified a functional single nucleotide polymorphism (SNP) c.1571 G>A situated in the 3' untranslated region (3'UTR) of *FADS2* in Chinese Holstein cows. Association analyses revealed that cows with the GG genotype had improved levels of delta-6 desaturase substrates (linoleic acid, C18:2n-6; $P < 0.001$) and decreased levels of desaturase products (gamma-linolenic acid, C18:3n-6; $P < 0.001$) in the milk, indicating a reduction in FADS2 expression or delta-6 desaturase activity caused by this polymorphism. Computer alignment demonstrated that c.1571G>A occurred within a potential miR-744 binding site. When the c.1571G allele was present, the luciferase activity of reporter constructs was significantly suppressed by miR-744, while no such effect was observed with the A allele. Overexpression of miR-744 in bovine mammary epithelial cells (with 1571GG genotype) downregulated FADS2 expression at both mRNA and protein levels. In contrast, inhibition of endogenous miR-744 with a specific inhibitor dramatically upregulated FADS2 expression. Taken together, these lines of evidence indicated that the c.1571A minor allele abolished the ability of miR-744 to bind FADS2, with a consequent increase in FADS2 expression levels and synthesis of omega-6 LC-PUFAs. Therefore, SNP c.1571 G>A could be used as a potential genetic marker in the selective breeding of cattle to increase beneficial FAs content in milk.

PE0340: Cattle

A Survey of RNA Methylation Within the Bovine AMPK Gene Family

Sean Thomas Quigley¹, Suraj Bhattarai¹, Thomas Buttolph¹, Brenda Murdoch², Jae Woo Kim³, Helen Yampara-Iquise³, Robert D. Schnabel⁴, Jeremy F. Taylor⁵, Sheryl White¹ and Stephanie McKay¹, (1)University of Vermont, Burlington, VT, (2)Department of Animal and Veterinary Science, University of Idaho, Moscow, ID, (3)University of Missouri, Columbia, MO, (4)Division of Animal Sciences, Informatics Institute, Columbia, MO, (5)Division of Animal Sciences, University of Missouri, Columbia, MO

Nucleic acids are able to undergo covalent modification in a variety of ways resulting in functional regulation at the epigenetic level. Understanding these epigenetic changes may provide insight into gene regulation, which can be applicable to any field within the life sciences that deal with DNA. Methylation of the carbon at the fifth position in the nucleic acid cytosine (5mC) within DNA is one of the most well documented epigenetic modifications. In eukaryotic organisms, 5mC is associated with down-regulation of mRNA transcription from the gene harboring methylated cytosines. Although cytosine methylation is well documented at the genetic level, little is known about it at a post-transcriptional level. Cytosine methylation within RNA has been observed in mRNA, tRNA, and rRNA however demonstrable functionality for this epitranscriptomic modification has yet to be elucidated. Stable and long-lived RNA molecules containing methylated cytosines have been described, possibly insinuating their role in RNA stability. The dearth of knowledge regarding the roles of post-transcriptional RNA modifications can be explained by the fact that 5mC is not easily detected by most sequencing methods and also that 5mC does not affect base pair binding, precluding the use of probing assays. In this study, a survey of mRNA methylation within the bovine AMPK gene family was conducted using skeletal muscle from six individual Charolais sired heifers from Angus x Simmental cows. We found that methylation of cytosines at the post-transcriptional level was detectable through the bisulfite treatment of RNA, followed by conversion to cDNA, PCR amplification and finally restriction digestion. These results provide a preliminary insight into the bovine AMPK epitranscriptome, and provide data

supporting post-transcriptional 5mC modification which suggests the need for further investigation in the relatively new sub-field of translational regulation. Finally, this study is the first survey of epitranscriptomic markers conducted within livestock species.

PO0341: Cattle

Genetic Signatures of African Indicine Cattle Reveals Candidate Modulators for Bovine Tuberculosis Resistance

Clementine Charton, Kwondo Kim, Taehyung Kwon and Heebal Kim, Seoul National University, Seoul, South Korea

Bovine tuberculosis (bTB) has been a major concern in cattle breeding for both its zoonotic potential and its economic impact. Though indicine (*Bos indicus*) cattle display greater resistance to bTB than taurine (*Bos taurus*) breeds, the genetic basis for this discrepancy has been scarcely explored. This study took advantage of whole genome sequencing data on three bTB resilient African indicine breeds and two bTB susceptible taurine breeds to decipher the genetic background underlying the bTB resilience discrepancy found between these two subspecies. Cross-population selection signature methods were used to identify immunity-related genes under selection that segregated between the two breed groups. As a complement, genes reported for modulating susceptibility to bTB were individually assessed for variants segregation between the two sub-species. A total of 20 candidate genes were proposed by combining the two aforementioned approaches. Two deleterious variants were found to segregate between indicine and taurine individuals within the *MARCO* gene, which has been shown to participate in the first step of mycobacterial signaling in macrophage. *MARCO* gene indeed codes for macrophage receptor with a collagenous structure, a class A scavenger preferentially used in binding one of the major mycobacterial cell wall lipid, cord factor. Deleterious variants were also found to segregate within three genes implicated in the signaling of mycobacterial antigens, *MAPK13*, *TLR6* and *TLR10*. The results presented here constitute a novel layer of information in understanding the genetic basis for bTB resilience and provide a stepping stone for deeper investigations.

PE0342: Cattle

Interactions of miRNA and mRNA in Response to Bovine Leukemia Virus Infection

Hao Ma, John D. Lippolis and Eduardo Casas, USDA ARS NADC, AMES, IA

Bovine leukemia virus (BLV) infection in cattle is omnipresent throughout the world, which causes significantly economical losses. The objective of this study was to evaluate cattle circulating miRNAs, mRNA profiles, and their interactions in response to BLV infection. Small RNA libraries were respectively sequenced from serum and white blood cells (WBC), which were sampled from seven BLV seropositive and seven seronegative cows. Transcriptome profiles were generated by sequencing RNA libraries prepared from the WBC. Bta-miR-206 and bta-miR-133a-3p were identified in serum to be differentially expressed between cows serologically positive or negative for BLV. While in WBC, bta-miR-335-3p, bta-miR-375, bta-novel-miR76-3p, and bta-novel-miR76-5p were differentially expressed between the two groups. RNA sequencing data analysis identified 69 differentially expressed transcripts (DETs), among which, 50 were targeted by the six miRNAs. The 50 DETs involved in various biological processes, including negative regulation of viral transcription, release from, and entry into host cells. The six miRNAs targeted 180 genes with predicted interaction energy smaller than -20 kcal/mol. Those genes mainly participated in RNA-dependent DNA biosynthetic process which may associate with viral proliferation. In addition, the bta-miR-206, bta-novel-miR76-3p, and bta-novel-miR76-5p interacted with BLV *rex*, *tax*, *gag*, and *env* genes by targeting their untranslated regions. Further studies of the miRNAs and the genes might reveal the molecular mechanisms of BLV infection and find possible ways to prevent the infection.

PO0343: Cattle

Expression of Viral microRNAs in Serum and White Blood Cells of Cows Exposed to Bovine Leukemia Virus

Eduardo Casas, Hao Ma and John D. Lippolis, USDA ARS NADC, AMES, IA

Bovine leukemia virus (BLV) causes abnormal immune function and immunosuppression. MicroRNAs (miRNAs) are known to be involved in gene expression and have been associated with stress and immune response, and tumor growth. The objective was to determine the expression of circulating miRNAs produced by the BLV in animals exposed to the virus. Sera from 14 animals were collected to establish IgG reactivity to BLV by ELISA, where seven animals were seropositive and seven were seronegative for BLV exposure. White blood cells (WBC) from each animal were also collected and miRNAs were extracted by sequencing from sera and WBC. The seropositive group had higher counts of BLV miRNAs when compared to seronegative group in sera and WBC. MiR-B1-3p, miR-B2-5p, miR-B4-3p, and miR-B5-5p were differentially expressed ($P < 0.00001$) in serum, while miR-B1-3p, miR-B1-5p, miR-B3, miR-B4-3p, miR-B4-5p, miR-B5-5p were differentially expressed ($P < 1.08 \times 10^{-9}$) in WBC, with an average of 7 log2 fold difference between the seropositive and the seronegative groups. MiR-B2-3p and miR-B2-5p were also differentially expressed in WBC ($P < 2.79 \times 10^{-17}$), with an average of 27 log2 fold difference between the seropositive and the seronegative groups. MiR-B1-5p and miR-B4-5p have been identified as being associated with response to stress and in the immune system process. There were 18 genes identified as being potential targets for miR-B1-5p, and 3 genes for miR-B4-5p. Several identified genes have been associated with leukemia development in humans and cattle. Differential expression of genes targeted by BLV miRNAs should be evaluated to determine their effect in BLV infection.

PE0344: Cattle

Profiling Host Transcriptome During Bovine Leukemia Viral Pathogenesis

Katy Kesler^{1,2}, Tasia M. Taxis², Tawnie Dejong², Kelly Sporer^{2,3} and Casey Droscha^{1,2}, (1)CentralStar Cooperative, Lansing, MI, (2)Michigan State University, East Lansing, MI, (3)CentralStar Cooperative, East Lansing, MI

Prevalence of bovine leukemia virus (BLV) has now surpassed 40% in US dairy cattle. Although BLV is often asymptomatic, progression results in persistent lymphocytosis leading to immunosuppression and decreased longevity. Current BLV diagnostics include detecting anti-BLV antibodies (ELISA) and determining proviral load (PVL) by quantifying viral DNA (qPCR) to screen and stratify disease status aiding in BLV management for producers. Yet, mechanisms of BLV transmission and pathogenesis remain ambiguous. The BLV provirus encodes anti-sense transcripts and microRNAs (miRNA) eliciting post-transcriptional modulation of host Bcell proliferation and maturation as a strategy for viral propagation. MicroRNAs are reliably detected in blood and have been shown to correlate with PVL ($r=0.928$), but miRNA functional relevance in BLV pathogenesis remains unclear. The purpose of this study was to elucidate the mechanisms of BLV progression, by developing miRNA profiles of animals in a progressive state of BLV infection (increasing PVL). Blood samples were obtained from a local cooperative dairy herd. BLV status was analyzed using ELISA and qPCR assays. Sequencing of mRNA and small non-coding RNA was performed using Illumina TruSeq library preparation technology. TaqMan qPCR assays were performed for validation of profiled miRNAs. The objective of this study is to describe transcriptional changes that occur during BLV pathogenesis to elucidate mechanisms behind BLV disease progression. Increased knowledge about the dynamics of BLV disease will enable improvement in diagnostic and prevention strategies to reduce BLV nationwide and increase the sustainability and profitability of the dairy industry.

PO0345: Cattle

NGS-Based Typing of MHC Class II Loci to Evaluate Resistance to Bovine Leukemia Virus in Commercial Dairy Herds

Chaelynne Lohr^{1,2}, Vickie Ruggiero², Kelly Sporer^{1,2}, Casey Droscha^{1,2} and Paul Bartlett², (1)CentralStar Cooperative, East Lansing, MI, (2)Michigan State University, East Lansing, MI

Bovine leukemia virus (BLV) infects approximately 50% of U.S. dairy cattle, of which 30% develop high proviral loads (PVL) and persistent lymphocytosis. This susceptible population have diminished immunocompetence, enabling the persistence of this retrovirus and other opportunistic pathogens within commercial herds, reducing the economic sustainability of the industry. The mammalian MHC locus, the bovine leukocyte antigen (BoLA) in cattle, is involved in presenting antigens to CD4+ or CD8+ lymphocytes. The extracellular portion of the class II BoLA DRB3 and DQA1 alleles, encoded within the second exon, is polymorphic with a diverse class of alleles. Our objective is to determine if specific BoLA alleles confer resistance or susceptibility to the progression of BLV and

define the BLV viral type for selected animals to assess viral type virulence. Multiplex sequencing of the second exon of the BoLA-DRB3 and DQA1 alleles were analyzed via end-to-end amplicon sequencing in a 2x250bp paired end format using a MiSeq v2 500 cycle flow cell (n = 384). To determine the BoLA haplotypes for each animal, trimmed reads were aligned to the reference haplotype sequences, obtained from EMBL-EBI. Viral typing was done by examining the BLV-envelope gene via Sanger sequencing. Analysis of phenotypically selected animals demonstrates that BLV resistance is associated with DRB3*0601, *0902, *1701 and DQA1*0204. Our data agrees with previously found associations between BLV resistance and DRB3*0902, *1701 and DQA1*0204.

PE0346: Cattle

Genomic Prediction of Resistance to *Babesia Bigemina* in Braford and Hereford Cattle

Andrea R. S. Romero¹, Henrique Nunes de Oliveira², Andre Vieira do Nascimento¹, Fernando F Cardoso³, Claudia C G Gomes⁴, Alexandre R. Caetano⁵ and Cedric Gondro⁶, (1)São Paulo State University, Jaboticabal, Brazil, (2)Sao Paulo State University, Jaboticabal, Brazil, (3)Embrapa Pecuária Sul, Bagé, Brazil, (4)Embrapa Pecuária Sul, Brasília, Brazil, (5)Embrapa Recursos Genéticos e Biotecnologia, Brasília, DF, Brazil, (6)Michigan State University, East Lansing, MI

Babesia bigemina is a protozoan parasite that causes bovine babesiosis which results in significant economic losses to the cattle industry. The objective of this work was to evaluate the accuracy of genomic prediction for infection levels of *Babesia bigemina* (ILBIG). Protozoa loadings for each animal were estimated by qPCR and used as proxy of infection levels. Data was also collected for infection levels by *Babesia bovis* (ILBOV) and tick counts (TC) for the same animals. Genetic and phenotypic information from 1485 Braford and 163 Hereford cattle was used. Animals were genotyped on 50k Illumina panels and imputed up to high density. After quality control 1638 animals and 701,570 SNPs remained. MTG2 software was used to adjust for fixed effects and estimate genomic values using univariate and multivariate models. Accuracy of estimates was evaluated by 100-fold cross-validation. Genomic prediction accuracy of ILBIG was low at 0.14 (± 0.06) with the univariate model. Accuracy increased to 0.19 (± 0.13) when ILBOV phenotypic information was included in the model. Tick counts however, had a detrimental effect on prediction accuracy of ILBIG, which had an accuracy of 0.03 (± 0.07) when they were included as a bivariate model and had an accuracy of just 0.12 (± 0.04) in a multivariate model that included ILBOV. Inclusion of ILBOV improves prediction accuracy of ILBIG, while TC values have the opposite effect and reduce it. Our results suggest that tick count measures are not useful to select for animals with increased resistance to *Babesia bigemina*. Grant FAPESP 2019/00412-3.

PO0347: Cattle

Divergent Pathogens and Genetic Predisposition Shape the Hepatic Transcriptome in Lactating Heifers After Intramammary Infection

Annika Heimes¹, Johanna Brodhagen¹, Rosemarie Weikard¹, Doreen Becker¹, Marie M. Meyerholz², Wolfram Petzl², Holm Zerbe², Martina Hoedemaker³, Laura Rohmeier², Hans-Joachim Schuberth⁴, Marion Schmicke⁵, Susanne Engelmann^{6,7} and **Christa Kuehn**^{1,8}, (1)Leibniz Institute for Farm Animal Biology (FBN), Dummerstorf, Germany, (2)Clinic for Ruminants with Ambulatory and Herd Health Services/ Ludwig-Maximilians-University Munich, Oberschleißheim, Germany, (3)Clinic for Cattle, University of Veterinary Medicine Hanover, Hanover, Germany, (4)Immunology Unit, University of Veterinary Medicine Hanover, Hanover, Germany, (5)Faculty of Natural Sciences III, Martin-Luther Universität Halle-Wittenberg, Halle, Germany, (6)Technical University Braunschweig, Institute for Microbiology, Braunschweig, Germany, (7)Helmholtz Centre for Infection Research, Microbial Proteomics, Braunschweig, Germany, (8)Agricultural and Environmental Faculty, University of Rostock, Rostock, Germany

In spite of respective breeding programs, bovine mastitis is still a major concern in the dairy industry and has implications for animal welfare and production and on antimicrobial drug use in livestock farming. Commonly, *Staphylococcus aureus* (*S. aureus*) induces a chronic and subclinical mastitis, while *Escherichia coli* (*E. coli*) infections usually result in an acute and clinical mastitis. However, clinical reports showed that there is a huge variability between cows in response to those pathogens. Thus, in our study we used lactating Holstein half-sib

heifers, which differed in the paternally inherited haplotype for a region on bovine chromosome 18, frequently described to exert major effects on health and longevity, and challenged the udder quarters with *S. aureus* or *E. coli*. In addition to monitoring the hepatic transcriptome with respect to intramammary pathogen challenge and haplotype, we compared the liver transcriptomes between pathogen-challenged individuals and untreated controls. In all animals intramammary challenge with either pathogen, *S. aureus* or *E. coli*, elicited systemic effects on transcriptional level. However, we observed pathogen-specific targeting strategies to bypass the host's innate immune system. While key components of the cytoskeleton were downregulated in the liver of *S. aureus* infected cows, individuals with intramammary *E. coli* challenge showed very strongly downregulated complement system and effects on metabolic hepatic pathways (e.g., lipid metabolism). Regarding genetic predisposition, *S. aureus* challenged cows with a haplotype predicted to be favorable for mastitis resistance displayed more activated immune genes and pathways compared to their half-sibs with the unfavorable paternally inherited haplotype.

PE0348: Cattle

Investigating the Influence of Exogenous Copper Supplementation on Copper Homeostasis in Beef Cattle Liver

Nicole M. Tillquist, Stephen J. Coleman and Terry E. Engle, Colorado State University, Fort Collins, CO

Copper (Cu) is an essential dietary trace element important for numerous biological functions including iron metabolism, lipid metabolism, hematopoiesis, ossification, fat deposition, cellular respiration, hair/wool pigmentation, immune system health, and central nervous system function. The primary organ responsible for Cu homeostasis is the liver. However, the majority of research examining hepatic Cu homeostasis has been conducted in rodent models. It is important therefore to understand bovine specific Cu metabolism to more efficiently supplement Cu. The objective of the current experiment was to investigate the influence of Cu concentrations on the expression of genes that are involved in Cu homeostasis in beef cattle. Liver samples were obtained immediately post mortem from healthy angus steers. Hepatocytes were isolated and cultured in treatments of 0ppm, 0.01ppm, 0.10ppm, 10.0ppm, and 100ppm Cu for 1 hour. Cells were collected and lysed in TRIzol™ total RNA isolated following a standard protocol. Quantitative RT-PCR was used to investigate the abundance of transcripts for proteins involved in Cu homeostasis in liver tissue and cultured hepatocytes. These genes were identified in a previous study within the department, and a literature review which identified Cu responsive proteins in sheep. The identified targets were: ALDH2, APOA1, ATOX1, ATP7A, ATP7B, BHMT, BLVRB, CA2, CCS, COX17, CTR1, ELN, GAPDH, GLUD1, GSS, LOXL1, PDIA3, SOD1, SOD3. The current analysis will investigate if there is a dose response in transcript abundance of Cu homeostatic proteins. Further investigation is needed to determine if liver Cu concentration influences Cu homeostatic protein function.

PO0349: Cattle

Genotype By Environment Interaction in Response to Cold Stress in a Composite Beef Cattle Breed

Sajjad Toghiani¹, El Hamidi Hay¹, Breno Fragomeni², Romdhane Rekaya³ and Andrew J. Roberts¹,
(1)USDA, Agricultural Research Service, Miles City, MT, (2)University of Connecticut, Storrs, CT,
(3)University of Georgia, Department of Animal and Dairy Science, Athens, GA

Cold stress influences the productivity and survivability of beef cattle in the Northern Great Plains region. The study objective is to evaluate the genotype-by-environment interaction impact due to cold stress on birth weight (BW) and weaning weight (WW). To assess the cold stress effect, a total cold load (TCL) calculated using the comprehensive climate index (CCI) and defined based on three temperature thresholds: less than -5°C (TCL5), -15°C (TCL15) and -25°C (TCL25). A total of 4,221 and 4,217 records for BW and WW respectively were used in both a univariate and a reaction norm model. As cold load increased, the direct heritability slightly increased in both BW and WW for TCL5 class. However, this heritability remained constant across TCL25 class. Contrarily, the maternal heritability of BW was constant with cold load increase in all TCL classes, although slightly increased maternal heritability of WW was observed for TCL5 and TCL15. The direct and maternal genetic correlation for BW and maternal genetic correlation for WW across different cold loads between all TCL classes were high ($r > 0.99$), whereas the lowest direct genetic correlations observed for WW were 0.88 for TCL5 and 0.85 for TCL15. The Spearman rank correlation between the EBV of top bulls ($n=79$) using univariate and reaction norm models across TCL classes showed some re-ranking in direct and maternal effects for both BW and WW particularly for TCL5 and TCL15. Overall, cold stress did not show a big impact on direct and maternal genetic effects of birth and weaning weight.

PE0350: Cattle

Profiling the Immune Epigenome across Global Cattle Breeds

Jessica R. Powell, The Roslin Institute, Edinburgh, United Kingdom

Understanding the variation between well and poorly adapted cattle breeds to local environments is essential for breeding cattle with improved climate and disease resistant phenotypes. While studies are beginning to characterise the genetic basis of cattle breed diversity, alternative mechanisms underlying breed-specific traits are largely unexplored. Changes at the chromatin level are of particular interest because of their potential role in disease. However, the tools and reference resources to study these changes in cattle are almost entirely lacking.

In this study, we have characterised the chromatin accessibility and DNA methylation landscapes genome-wide of seven immune cell types across three diverse cattle breeds. Holstein Friesian, N'Dama and Nelore cattle were selected to represent the European taurine, African taurine and indicine cattle lineages respectively. Gene expression data for Holstein Friesian cattle have also been generated.

We analysed the variation across immune cell types and breeds and found more distinct breed differences between cell types of the adaptive than the innate immune system. These unique cell type profiles enable the accurate deconvolution of complex cellular mixtures using digital cytometry approaches. Finally, we show distinct sub-categories of CpG islands based on their chromatin and methylation profiles that discriminate between classes of distal and gene proximal islands linked to discrete transcriptional states.

These data provide an extensive resource to help exploit the diversity among cattle breeds and improve cattle productivity for farming communities in low and middle income countries.

PO0351: Cattle

Genetic Contribution of Admixture to Adaptation of Indigenous African Cattle

Kwondo Kim¹, Taehyung Kwon¹, Olivier Hanotte² and Hee-bal Kim¹, (1)Seoul National University, Seoul, South Korea, (2)School of Life Sciences, University of Nottingham, Nottingham, United Kingdom

Across African human societies, cattle play a crucial role as a source of animal proteins, draught power and wealth. The majority of African cattle are today indicine x taurine admixed populations with high genome diversities, a legacy of crossing between African taurine and multiple-introduced indicine. They are present across all agro-ecologies, from the driest to the most humid ones, with their successful adaptations to local environmental conditions. We present here several evidences that such success of African cattle pastoralism was driven by the indicine x taurine admixture. We analyzed whole genome sequences of 162 indigenous African cattle, representing 15 populations (taurine, sanga, zenga, and zebu) and showed that the main taurine x indicine admixture events in African humped cattle date back to around 650 years ago. We then reconstructed the local ancestries of African humped cattle, and identified signature of positive selection. The selective sweeps with an excess of indicine as well as taurine ancestry include annotated genes underlying adaptive traits of unadmixed indicine or taurine cattle. Analyzing two African taurine populations admixed with indicine, we also identified a conserved haplotype commonly shared within African taurine, clearly distinct from Eurasian taurine and Asian indicine haplotypes. This unique haplotype is located upstream *CARD11*, a gene previously reported to be linked to trypanotolerance. It represents an example of selection signatures specific to African taurine cattle. Our findings suggest that a combination of different genetic resources shaped the genetic background of African cattle, and contributed the rapid dispersion of cattle across diverse African agro-ecologies.

PE0352: Cattle

Towards a Complete Genome Characterization of African Indigenous Cattle

Abdulfatai Tijjani^{1,2}, Karen Marshall^{2,3}, Endashaw T. Assegidaw⁴, Jian-Lin Han⁵, Hee-bal Kim⁶, Joram M. Mwacharo⁷ and Olivier Hanotte^{8,9}, (1)International Livestock Research Institute (ILRI), Addis Ababa, Ethiopia, (2)Center for Tropical Livestock Genetics and Health (CTLGH), Midlothian, United Kingdom, (3)International Livestock Research Institute (ILRI), Nairobi, Kenya, (4)Addis Ababa University, CNS,

MCMB Department, Addis Ababa, Ethiopia, (5)CAAS-ILRI Joint Lab, Inst. of Anim. Sci., Beijing, China, (6)Seoul National University, Seoul, South Korea, (7)International Centre for Agricultural Research in Dry Areas, Addis Ababa, Ethiopia, (8)School of Life Sciences, University of Nottingham, Nottingham, United Kingdom, (9)LiveGene – CTLGH, International Livestock Research Institute, Addis Ababa, Ethiopia

Genetic and genomic variations in African indigenous livestock populations remain huge resources, yet to be tapped. These resources are prerequisite for the complete characterization of local breeds towards genetic improvement contributing to food security and poverty reduction on the continent. The International Livestock Research Institute (ILRI) is leading an unprecedented research effort aiming at harnessing the genetic resources of indigenous cattle breeds across Africa. This is being achieved through an extensive collaboration of African national partners on ILRI's LiveGene-Genetic reference resource for African cattle (GRRFAC) project, supported by the University of Nottingham (UK), the ILRI - CAAS Joint Laboratory on Livestock and Forage Genetic Resources (China), Seoul National University (South Korea) and the Center for Tropical Livestock Genetics and Health (CTLGH) program based at the University of Edinburgh and Scotland Rural College (Roslin Institute, UK). So far, samples have been obtained from around 40 African indigenous cattle breeds of zebu, taurine and admixed origins and comprising of over 1,500 individuals. High coverage (up to 30x) of whole genome re-sequencing of more than 25 breeds consisting of over 400 samples have now been completed. We aim to catalogue the entire functional diversity, including SNPs and structural variations, of African cattle. Our current objectives include the identifications of the most informative SNPs for designing the African cattle reference SNP genotyping arrays to be applied to breeding improvement programs and the genomic regions underpinning adaptation and productivity, as well as the capacity building among African collaborating scientists through training in bioinformatics. Future activities may include *de novo* genome sequencing of African cattle breeds and a transcriptome catalogue of gene expression, paving the way to the pan-genome analyses of African cattle.

PO0353: Cattle

Validation of the ISAG Bovine SNP Panel for Parentage Verification in South African Bonsmara and Drakensberger Cattle

Yandisiwe P. Sanarana^{1,2}, Azwihangwisi Maiwashe¹, Cuthbert B. Banga¹, Donagh P. Berry^{2,3} and Este van Marle-Köster², (1)Agricultural Research Council, Pretoria, South Africa, (2)University of Pretoria, Pretoria, South Africa, (3)Teagasc, Fermoy, Ireland

Single nucleotide polymorphisms (SNPs) have replaced microsatellite markers in most applications. A bovine panel consisting of 200 markers has been recommended by the International Society for Animal Genetics (ISAG) for parentage verification. However, the applicability of the panel in South African (SA) populations, which are predominately *Bos taurus* and *Bos indicus* cattle and their crosses, has not been evaluated. In this study, the information content and usefulness of the panel in SA Bonsmara (BON) and Drakensberger (DRB) were investigated. Genotypes for the 185 ISAG SNPs were available for 45 BON and 74 DRB sire-offspring pairs. Exclusion was considered whenever the genotype of the sire was discordant with the genotype of the offspring for more than one of the SNP. Twenty-one of the 185 ISAG SNPs were either monomorphic, had a poor call rate, or clustering score. Although the mean minor allele frequency of the ISAG SNP was 0.331 and 0.359 in the BON and DRB, respectively, there was no difference in the power of probability of parentage exclusion (PE) (99.46%) between the two breeds. On average, the ISAG panel confirmed parentages of 32 and 50 animals of the true sire-offspring pairs BON and DRB to exist, respectively, from genotyped 150,000 SNPs on all animals. This means that the panel failed to confirm parentages of 28.9 and 32.4% for the BON and DRB, respectively. As shown in the study, more informative SNP markers are necessary for accurate parentage testing in SA cattle.

PE0354: Cattle

Assessment of Genetic Structure and Admixture Analysis of Smallholder Dairy Cattle Herds in South Africa Using SNP Markers

Cuthbert B. Banga^{1,2}, Maake Mphapantsi^{2,3}, Obert Tada³, **Yandisiwe P. Sanarana**² and Farai C. Muchadeyi⁴, (1)Tshwane University of Technology, Pretoria, South Africa, (2)Agricultural Research

Council, Pretoria, South Africa, (3)University of Limpopo, Polokwane, South Africa, (4)ARC-Biotechnology Platform, Pretoria, South Africa

Poor animal productivity, primarily due to a lack of genetic improvement programs, is a major concern for the smallholder dairy sector in South Africa. There is limited information for developing such programs, such as current breeding practices and genetic makeup of cattle used in this production system. This study was carried out to evaluate the levels of genetic diversity and population structure in South African smallholder dairy (SHD) herds using SNP markers. A total of 192 animals, randomly sampled from South African smallholder dairy herds, were genotyped using the GeneSeek® Genomic Profiler (GGP) 150K-bead chip. About 846 animals from four commercial breeds, Ayrshire (n=200), Holstein (n=231), Jersey (n=224) and a local breed Nguni (n=209), genotyped with the BovineSNP50 chip, were used as reference populations. After quality control, the data was analyzed to assess genetic diversity and population structure using Plink, GCTA and Admixture software. Measures of within breed diversity across the five populations were comparable (ranges: F_{is} -0,02 - 0,02; H_o 0,39 - 0,40; H_E 0,39 - 0,40; MAF 0,3 – 0,31), with the lowest diversity and highest genomic inbreeding being observed in the smallholder population. Five distinct populations were observed, with most individuals from the smallholder population being closely related to Holstein and Jersey. This implies widespread crossbreeding in the smallholder herds, mainly involving Holstein and Jersey breeds, with limited use of the Nguni breed. These results provide a useful insight into the genetic structure and prevailing breeding practices on smallholder dairy herds.

PO0355: Cattle

Mito-Nuclear Incompatibility: The Essential Drive behind the Absence of Zebu Mitochondria in Admixed African Cattle

Taehyung Kwon¹, Kwondo Kim¹, Olivier Hanotte² and Heebal Kim¹, (1)Seoul National University, Seoul, South Korea, (2)LiveGene – CTLGH, International Livestock Research Institute, Addis Ababa, Ethiopia

Despite the zebu-aurine admixed nature of the African cattle, only aurine mitochondrial haplotypes are found on the African continent. Male-mediated zebu introgression is often considered an explanation for this contrast, although it has not been thoroughly investigated with empirical genomic data. Here, we propose mito-nuclear incompatibility, a well-known barrier to interspecies hybridization, is one important culprit. We identified that the absence of zebu haplotype in full mitochondrial sequences contrasts the prevalence of zebu genetic contribution in 162 whole genome sequences of African cattle we here characterize. Based on the demographic simulation, we conclude that this contrast in African humped cattle necessitates selections based on both zebu male preference and mito-nuclear incompatibility. Furthermore, we identify selection signatures of the mito-nuclear incompatibility on the genome of African zebu cattle, which supports the impact of adaptive introgression against aurine mitochondria. Our findings expand the knowledge of the evolutionary history of African cattle and suggest a novel perspective on cattle evolution.

PE0356: Cattle

Periparturient Period Affects Global Gene Expression in Holstein Friesian Cow Neutrophils

Emmanuel Asiamah¹, Kingsley Ekwemalor², Sarah Adjei-Fremah², Bertha Osei³ and Mulumebet Worku², (1)University of Arkansas, Pine Bluff, AR, (2)North Carolina Agricultural and Technical State University, Greensboro, NC, (3)Oklahoma Medical research Facility, OKLAHOIMA CITY, OK

The periparturient period is defined as the period 3 weeks before parturition (calving) and 3 weeks after parturition. This period is characterized by dysfunctional immune responses as a result of the impairment of the morphology and functions of neutrophils. These changes are associated with altered gene expression in the neutrophils which leads to immunosuppression. A plethora of research is emerging to elucidate the molecular mechanisms underlying the impaired immune function in dairy cows during the periparturient period

This study aimed at evaluating the global gene expression of blood-derived neutrophils from periparturient cows. Blood was collected from Holstein Friesian periparturient cows (N=3) at -14 d relative to expected calving date and 7 d relative to the actual calving date. Neutrophils were isolated and subsequently used for transcriptional profiling using the Agilent bovine (v2) 4 × 44 K array. Calculation of fold change in gene expression and pathway analysis was conducted using the GeneSpring GX software 13.0.

The results showed that 249 genes were differentially expressed ($FC \geq 2$, $p < 0.05$); 162 were upregulated post-calving 87 of these were downregulated. Genes that code for pro-inflammatory receptors (CD58, GLRX3), chemokines (CMKLR1), and transcriptional regulation (MTA) were upregulated. Concurrently, genes that code for cellular adhesion and migration (ADRM1 and THY1), and immune induction (CATHL2) and homeostasis were downregulated gene. Pathway analysis revealed that 118 pathways are affected in bovine neutrophils during the periparturient period ($p < 0.05$). These pathways included the Wnt signaling, one carbon Metabolism, TLR, inflammation response, Oxidative Stress, T-Cell Receptor signaling, adipogenesis, and MAPK Signaling Pathways. This new knowledge generated about altered gene expression in neutrophils of periparturient dairy cows will ultimately be used for the development of novel management strategies to combat immunosuppression and disease susceptibility during this stage in the dairy cow.

PO0357: Cattle

Identification of Potential Biomarkers for Disease in the Microbiome of the Peripartum Reproductive Tract and Colostrum of Multiparous Holstein Cows.

Connor E. Owens, Haley G. Huffard, Alexandra Irma Nin-Velez, A. Jane Duncan, Katharine F. Knowlton and Rebecca R. Cockrum, Virginia Polytechnic Institute and State University, Blacksburg, VA

The objective of this study was to further characterize the microbiome of the dam's reproductive tract and colostrum at calving and identify potential biomarkers for post-partum disease. Swabs of the posterior vagina were collected 24 h before calving (V; $n = 4$), colostrum was collected within 1 h after calving (C; $n = 6$), and placenta was collected within 6 h after calving (P; $n = 5$). 16S libraries of extracted DNA were created and sequenced via Illumina MiSeq V3, 300 PE. Operational Taxonomic Unit (OTU) clustering was performed in CLC Genomics Workbench (ver. 11.0.1) using 97% Greengenes reference database. PERMANOVA analyses and differential abundance analyses were performed with the main effects of sample location or calf sex. The dominant phylum in all locations was *Proteobacteria* (V = 58%, C = 96%, P = 48%). The vaginal microbiome was different from the microbiome of colostrum ($P = 0.014$) and placenta ($P = 0.048$), but colostrum and placenta did not differ ($P = 0.266$). There was no difference in the microbiomes based on calf sex ($P = 0.855$). Of the 47 OTUs that differed between locations, *g Streptococcus* had the greatest difference, being more abundant in the vagina than colostrum or placenta (\log_2 fold-change = 15.61). The most abundant OTUs in placenta were *g Acinetobacter*, *g Corynebacterium*, *g 5-7N15*, and *g Stenotrophomonas*, while in colostrum were *g Stenotrophomonas*, *f Pseudomonadaceae*, and *g Yersinia*. Further investigation of the relationship with these opportunistic pathogen OTUs in placenta and colostrum and post-partum diseases could provide biomarkers to prevent incidence of disease.

PE0358: Cattle

Sequence and Assembly of the Holstein Y Chromosome

Liushuai Hua, The Pennsylvania State University, University Park, PA, Gonzalo Rincon, Zoetis, Kalamazoo, MI and **Wansheng Liu**, Department of Animal Science, Penn State University, University Park, PA

The Holstein male lineages rapidly reduced since the 1950s because of extensive use of artificial insemination in dairy cattle. All Holstein bulls today can be traced to one of two male lineages, Round Oak Rag Apple Elevation (Elevation) and Pawnee Farm Arlinda Chief (Chief). The objective of this work was to sequence, assemble and annotate the Holstein Y chromosome. DNA from a bull of the Elevation lineage was sequenced with multiple data types, including the illumina paired-end (PE), illumina mate pair (MP), PacBio long reads (PB), and Dovetail Chicago reads (Hi-C). A draft assembly of the Holstein Y chromosome was built and analyzed by combining a set of bioinformatics tools. The initial Y-linked contigs was assembled from the PE reads, resulting in a total length of 17.3 Mb. The contigs were scaffolded with MP reads, and gaps of the scaffolds were filled with PB reads, and further improved by Hi-C reads. The final Holstein Y assembly had a total length of 24.3 Mb, including 7208 contigs with N50 7618 bp. Compared to the Hereford Y sequence, the X degenerate region in the Holstein Y draft assembly was in a relatively high quality, where all 12 Y-linked single copy genes were annotated. However, the Y ampliconic region in the draft was incomplete because of the highly repetitive sequences. Sequence variations were found in three genes, including RBMY, UBE1Y and USP9Y. The copy numbers of all Y-linked genes were

estimated. Results from this work provides sequence data and a framework for studying Y-gene functions and evolution of the bovine and other mammalian Y chromosome.

PO0359: Cattle

A Structural Variant Is Associated with Tetradysmelia in Holstein Friesian Cattle

Doreen Becker, Rosemarie Weikard and Christa Kuehn, Leibniz Institute for Farm Animal Biology (FBN), Dummerstorf, Germany

Tetradysmelia is a rare congenital disorder characterized by a severe reduction of all limb parts distal of scapula and pelvic girdle. Besides being caused by non-genetic factors, tetradysmelia can have a genetic etiology. Herein, we have identified a structural variant in a Holstein Friesian backcross family with six stillborn calves displaying tetradysmelia. Pathological examination of available cases showed all affected animals having a rather uniform and severe dysmelia of all four limbs. Pedigree analysis suggested an autosomal recessive inheritance. Three cases and 14 related controls were genotyped using a bovine Illumina 50k BeadChip. By homozygosity mapping we identified a 10.54 Mb spanning homozygous region on bovine chromosome (BTA) 14 exclusively shared by the affected offspring. To further fine-map the candidate region, the whole genome of the cases' sire was sequenced. Subsequent data filtering against 3,102 control genomes of the 1000 Bull Genomes Project did not identify a candidate variant. Consequently, we performed a structural variation analysis and identified a 50 kb deletion in the candidate region on BTA 14. Screening of all available family samples confirmed perfect segregation of the deletion with tetradysmelia. The deletion affects a gene that enhances Wnt signaling and is involved in a broad range of developmental processes. Only recently, several mutations in the human homolog causing a similar phenotype were described. Therefore, we propose the 50 kb deletion on BTA14 as causative for bovine tetradysmelia. To our knowledge, this is the first reported mutation for tetradysmelia in a large animal model.

PE0360: Cattle

Recent Selection Footprints Associated with Aggressiveness in Bovines of the Lidia Breed

Paulina G. Eusebi, Velogen A.S.L / Universidad Complutense de Madrid, MADRID, Spain

Aggressiveness is a primitive behavior whose molecular mechanisms are similar and common among the subphylum vertebrata. In humans, a primary goal to study the molecular basis underlying aggression studies is to determine the neurobehavioral factors triggering violence. Several animal models as murines have been used to study agonistic responses, however aggression must be artificially induced, as rodents are not selected for these traits; developing as a consequence more intense aggressive reactions. Conversely, the Lidia cattle is a population selected since the 18th century to display aggressiveness by a set of behavioral traits with significant heritability values; this intensive selection may have driven specific allele frequency shifts. A previous analysis across the autosomes revealed long-term selection regions including genes involved in behavioral development. This study focuses on mapping recent signatures of selection at chromosome X associated with aggressiveness, by comparing Lidia cattle samples with two non-specialized Spanish breeds showing tamed behavior. The most significant markers analysed peak around the monoamine oxidase A (*MAOA*) gene, and thus the association of three functionally important regions located near the promoter of this gene were further investigated. A polymorphism consisting of a variable number of tandem repeats of the nucleotide "C" (BTX:105,462,494) was detected displaying lower number of repetitions in the Lidia breed when compared with the tamed breeds. An *in silico* analysis predicted that the g.105,462,494delinsC variant may code for the Sp1 binding motif, one of the major transcription factors controlling the core promoter and expression of the *MAOA* gene in humans.

PO0361: Cattle

Genetic Evaluation of the Akaushi Breed in the United States

Hanna A Wackel¹, Jason Rowntree¹, Dajeong Lim² and Cedric Gondro¹, (1)Michigan State University, East Lansing, MI, (2)National Institute of Animal Science, Jeonju, South Korea

The Japanese brown cattle, known as Akaushi, are well regarded due to their excellent marbling ability. The breed was imported into the United States in the 1990's when 10 animals were shipped from Japan to Texas. The population has steadily grown to around 10,000 head currently in the US. Population parameters and their

relationship to other cattle breeds have not yet been explored for this breed. For this study we genotyped 43 highly-representative Akaushi sires on the Illumina 50k SNP chip and characterized the Akaushi's main population parameters – inbreeding, heterozygosity, opposing homozygotes, linkage disequilibrium, effective population size and also compared them to other Asian and European cattle breeds through principal component analyses. Results showed that Akaushi cattle are closely related to each other due to high linkage disequilibrium and some animals had large inbreeding coefficients. Heterozygosity levels are lower than other breeds but there is still considerable variation. In relation to other breeds, the Akaushi are distantly related from European breeds and are most closely related to the South Korean Hanwoo, North Korean Chosun, and Chinese Yeonbyun. Even though the breed is generally regarded as being from Japan, it is less related to the Japanese Black Wagyu and maintains a close genetic relationship to the ancestral Korean peninsula breeds from which it derived.

PE0362: Cattle

Accuracy of Genomic Breeding Values for Ultrasound Carcass Traits Using Single-Step GBLUP in Montana Tropical® Composite Beef Cattle

Lais Grigoletto^{1,2}, Luiz Brito¹, Hinayah R. Oliveira¹, Joanir P. Eler², Fernando Baldi³ and Jose B. S. Ferraz², (1)Purdue University, West Lafayette, IN, (2)University of Sao Paulo/College of Animal Science and Food Engineer/FZEA, Pirassununga, Brazil, (3)São Paulo State University, Jaboticabal, Brazil

The incorporation of genomic information in genetic evaluations has substantially contributed to increase the accuracy of breeding values. However, genomic evaluations in Composite populations is still challenging. For instance, the Montana Tropical Composite beef cattle population was developed in Brazil based on the crossing among four different biological types derived from various *Bos taurus* and *Bos indicus* breeds. In this study, the traditional BLUP (based on pedigree-based relationship matrix) and a single-step genomic BLUP (ssGBLUP) models were compared to estimate breeding values for ultrasound carcass traits in the Montana Tropical animals. The phenotypic dataset included ~9,000 records for *Longissimus* muscle area (LMA), backfat thickness (BFT), rump fat thickness (RFT) and marbling score (MARB). The pedigree file included 28,654 animals, from which 1,900 were genotyped using a 30K, 35K, 50K or HD SNP panel and imputed to a common 50K SNP panel containing 50,980 SNPs. Genotyped animals were divided based on their birth year into training (n=1,452) and validation (n=448) populations. Accuracy was assessed based on the Pearson correlation coefficient between estimated breeding values (EBVs) and genomic EBVs (GEBVs). Prediction bias was assessed based on the linear regression coefficient estimated between EBVs and GEVs. The ssGBLUP model increased the accuracy of GEBVs (average = 0.61) across traits by 27% compared to EBVs. Furthermore, the ssGEBVs were less biased (average = 0.03) compared to EBVs. Our findings indicated a 27% increase in accuracy of breeding values for ultrasound-based carcass traits by incorporating genomic information through the single-step procedure.

PO0363: Cattle

Association of PLAG1 and SCD on Growth Performance, Carcass Characteristics, and Fatty Acid Composition in Hanwoo

Kim Margarette C. Nogoy¹, Hyeona Lee², Yan Zhang¹, Jia Yu¹, Ye Hyun Lee¹, Jin Ho Kim³, Hyun-A Seong² and Seong Ho Choi¹, (1)Department of Animal Science, Chungbuk National University, Cheongju, South Korea, (2)Department of Biochemistry, Chungbuk National University, Cheongju, South Korea, (3)Chung-Dam EL Hospital, Cheongju, South Korea

Association of pleomorphic adenoma gene 1 (PLAG1) and stearoyl CoA desaturase (SCD) genes on growth performance, carcass characteristics, and fatty acid composition in the native Korean cattle (Hanwoo) were evaluated. Muscle samples were immediately collected from newly slaughtered Hanwoo steers (n=42). Single-nucleotide polymorphisms (SNPs) in the PLAG1 and SCD genes using polymerase chain reactions-single strand conformation polymorphism were conducted. SNP g.25,003,337 (3'UTR) and three genotypes: CC, CG and GG (374bp, 374/237,135bp, 237/136bp) were detected in PLAG1. SNP g.10329 C>T located in exon 5 and three genotypes: TT, CT, and CC (494bp, 494/280/214bp, 280/214bp) were detected in SCD. In both genes, genotypes had no significant association to growth performance and carcass characteristics. However, it should be noted that Hanwoo with PLAG1 carrying CG genotype tended to increase final body weight, slaughter weight and dressed weight (p=0.053, p=0.53, p=0.067). PLAG1 carrying CG genotype in Hanwoo also showed significantly high

backfat thickness ($p < 0.049$). Both genes had no significant association to carcass quality traits such as intramuscular fat, meat and fat color, texture and meat quality grade except that SCD carrying CT genotype had high maturity trait. Meat quality traits and fatty acid composition were also not significantly associated to any of genotypes carried by PLAG1 and SCD. These results could be due to low population number of Hanwoo steers used and the alteration of feed diets given to steers during the late fattening period. Continuous research is being conducted to fully investigate the association of SCD and especially PLAG1 genes as DNA markers for improving meat quality traits and growth in Hanwoo.

Keywords: PLAG1, SCD, carcass characteristics, growth performance, Hanwoo

PE0364: Cattle

Assessment of Genomic Prediction Accuracy for Meat Quality Traits using Various SNP Densities in Hanwoo Cattle

Mohammed Negash Bedhane¹, Julius H van der Werf¹, Nasir Moghaddar¹, Dajeong Lim², Byung-Ho Park³, Mi Na Park⁴, Roh Seung Hee⁴, Sara de las Heras-Saldana⁵ and Samuel Clark⁵, (1)School of Environmental & Rural Science, Armidale, NSW, Australia, (2)National Institute of Animal Science, Suwon, Korea, Republic of (South), (3)National Institute of Animal Science, RDA, Wanju, South Korea, (4)Animal Genetic Improvement Division, National Institute of Animal Science, Rural Development Administration, South Korea, (5)School of Environmental & Rural Science, Australia

The availability of genome-wide single nucleotide polymorphism (SNP) panels has enabled the implementation of genomic prediction in many livestock species. Genomic prediction is widely applied to estimate genomic breeding values (GBV) since it was first proposed by Meuwissen in 2001. In beef cattle, genomic prediction has promising benefits for the improvement of carcass traits such as meat quality, because estimated breeding values can be obtained without sacrificing the selection candidates. The accuracy of genomic prediction mainly depends on the size and the diversity of the reference population, heritability of trait and the linkage disequilibrium between SNP and QTL. With whole-genome sequence (WGS) data, it is assumed that the causal mutations responsible for trait variation are included in the data, and therefore, the accuracy of prediction is expected to be improved compared to common SNP panels. The objective of this study was to examine the effect of various SNP densities (50K, HD and WGS) on genomic prediction accuracy for meat quality traits in Hanwoo beef cattle. Genomic and phenotypic data from 2,110 animals were used to predict genomic estimated breeding values (GBV) for marbling score (MS), meat texture (MT) and meat colour (MC). The 2110 Hanwoo steers were divided into 10 folds cross-validation using random sampling of individuals. Each of the fold ($n=211$, 10%) was used as validation dataset whereas the rest of the animals ($n=1899$, 90%) were used as a reference population. The WGS data (~15 million SNPs) was imputed from the 50K SNP chip to 777K, followed by an imputation step up to the whole-genome sequence level. The accuracy of imputation for WGS was on average 78% for SNPs with a MAF >0.01 . The genomic best linear unbiased prediction model was used to predict the GBV for each trait fitting either of the genomic relationship matrices from the 50k, HD, and WGS data. Then the accuracy of GBV was assessed using the Pearson's correlation between GBV and corrected phenotypic value divided by the square root of heritability. The estimated genomic prediction accuracies for MS, MT, and MC were 0.45, 0.39 and 0.29, respectively using either WGS or HD SNP panel. However, the 50K SNP panel yielded slightly higher prediction accuracies for MS (0.46) and MC (0.31) traits than the other panels. The prediction accuracy of MT (0.39) was similar for all SNP densities. The result showed that the high-density SNPs (WGS and HD) did not improve the genomic prediction accuracy for all studied traits.

P00365: Cattle

Analysis of Copy Number Variation in Jersey Dairy Cattle Using Whole Genome Sequencing

Beth M Lett, University of Wisconsin Madison, Madison, WI, Brian W. Kirkpatrick, University of Wisconsin-Madison, Madison, WI and Xian Qiao, INNER MONGOLIA AGRI UNIVERSITY, HOHHOT, China

Copy number variation (CNV) plays a role in disease resistance/susceptibility, fertility, milk production, and phenotypic variation. The objective of this study was to 1) identify CNVs in Jersey cattle using whole genome sequencing (WGS) and 2) validate predicted CNVs located within genes. WGS Illumina paired-end short reads at

15x coverage were generated for twenty Jersey AI sires and four USDA-MARC twinner sires, the latter representing a genetically diverse, outbred population. All sequences were aligned to the ARS-UCD1.2 reference genome assembly using BWA-MEM. CNV discovery utilized four methods: CNVnator, DELLY, LUMPY-Single sample, and LUMPY-Population. Consensus CNVs were identified by comparing results across the four methods within and among all samples (n=1,269 CNVs), Jersey only samples (n=740 CNVs), and non-Jersey only samples (n=59 CNVs). Putative Jersey-only CNVs were considered for validation after restriction to CNVs in functional gene regions (n = 86) and CNVs of low frequency outside functional regions (n=41). Validation consisted initially of comparing CNV locations with those reported in the database of genome variants (DGVa) from Ensembl and other published reports (included: 66/86 and 22/41). Putative CNVs not in DGVa or publications were validated using PCR to amplify deletion (duplication) or non-deletion (non-duplication) products. 3/20 (17 in progress) were validated by PCR. Among these validated, functional region CNVs, six were found to have no deletion homozygotes in our preliminary sample of 20 Jersey sires. Further investigation is warranted to determine if absence of homozygotes remains true in a larger sample of Jersey animals, suggesting the possibility of embryonic lethality.

PE0366: Cattle

Utilization of Microsatellites to Assist in Breed Management Strategy of the Lynch Lineback Cattle Within Canada

Pamela Walker and Carl Lessard, Agriculture and Agri-Food Canada, Saskatoon, SK, Canada

The Lynch Lineback is a landrace breed of cattle developed by the Lynch family. Subsequently, breeding management was closed for nearly a century. Recently, the breed was opened to other producers. However, the number of cattle available is low and the breed status is critical. Lynch Lineback producers want to preserve the genetic diversity this rare breed. The goal of this study was to estimate individual heterozygosity values, which may assist in breed management strategies to improve the Lynch Lineback breed. DNA was extracted from hairs obtained from 55 Lynch Lineback, representing the cattle produced in Ontario. A total of 23 microsatellites was used to genotype the individuals. GenAlEx 6.5 was used to estimate overall population diversity and to test the breed purity. GENHET, a function written for the program R, was used to determine 5 different parameters of heterozygosity. Our results confirmed that all 55 cattle tested assigned to the Lynch Lineback. Expected heterozygosity (H_e) was 0.450 ± 0.029 , which is consistent with low population numbers. In addition, the mean proportion of heterozygous loci in an individual (PHt) was 45%, with the lowest PHt of 21% and highest PHt of 68%. A total of 29 out of 55 animals had individual PHt values above the mean PHt. Thus, selection of breeding pairs using known pedigree data to ensure limited relatedness and selecting animals of higher PHt scores will assist in the preservation of genetic diversity within the Lynch Lineback, securing the retention of breed fitness in upcoming generations.

PO0367: Cattle

Genome-Wide Association Study for Hair Length in Brangus Heifers

Kaitlyn M Sarlo Davila, Department of Animal Sciences, University of Florida, Gainesville, FL

Thermal stress limits beef cattle production and results in a loss of \$370 million in the U.S. annually due to reduced animal performance. A shorter hair coat is a key thermoregulative adaptation that allows cattle to lose heat more efficiently through conductive, convective, and evaporative cooling at the hair-skin interface. The objective of this study was to identify genetic variants associated with the length of the topcoat and undercoat of cattle. Hair samples were collected from the shoulder, 4 inches down from the spine from 1456 heifers in 2016 and 2017. ImageJ software was used to measure hair length. The length of the topcoat and undercoat were evaluated for each individual by averaging five long and five short hairs, respectively. DNA was extracted from blood samples and genotyped with the Bovine GGP F250 array. After quality control, 109,538 SNP were available for association analyses using the univariate procedures of GEMMA that fitted the genomic relationship matrix to account for the genetic covariance among animals. To correct for multiple tests, the Benjamini-Hochberg false discovery rate was constrained to 0.2. Four SNP in the PRLR gene were significantly associated with topcoat length. The SLICK mutation in PRLR has previously been demonstrated to significantly impact hair length in cattle. Seven SNP in the PCCA gene were significantly associated with undercoat length. PCCA belongs to the biotin transport and metabolism pathway. Biotin deficiency has been reported to cause hair loss. These genetic variants may contribute to a shorter hair coat and more thermotolerant animals.

PE0368: Cattle

Beyond the Genome: What Having a Platinum Quality Brahman (*Bos indicus*) Genome Can Teach Us about What Makes a Useful Reference Genome

Elizabeth Ross, Queensland Alliance For Agriculture and Food Innovation - University of Queensland, Brisbane, QLD, Australia

Brahman cattle are a member of the Zebu (*Bos indicus*) subspecies of cattle. Despite their known genetic distance from taurine cattle genomic research has historically utilised the taurine reference genome, as it was the only available option. With the rapid reduction in price of long read sequencing this is no longer the case. We have built a platinum quality reference genome for Brahman cattle using 195GB of PacBio data, Hi-C and Chicago scaffolding and 160GB of Illumina short reads for polishing. The resulting assembly contains all chromosomes in single scaffolds, with only 330 gaps. However it is not only the accuracy and completeness of a reference genome that makes it a valuable research tool. Additional long read Oxford nanopore data from 14 Brahman including the offspring of the reference animals allows us to investigate structural variants in Brahman cattle. Additionally, the interaction of topologically associating domains (TADs) with expression data from blood, and full length transcript Isoseq data from 11 of the reference animal's tissues allow the investigation variation not only of the genome, but of the genome function.

PO0369: Cattle

Genes Regulating Calcium Availability and Utilization in Angus Steers May be Useful in Identifying Cattle with Reduced Susceptibility to Pulmonary Hypertension in High Altitude Beef Production Systems

Natalie F. Crawford¹, **Kaysie J. Jennings**¹, Angela Cánovas², Juan F. Medrano³, Alma Islas-Trejo⁴, R. Dale Brown⁵, Stephen J. Coleman¹, Scott E. Speidel¹, R. Mark Enns¹, Kurt R. Stenmark⁵ and Milt G. Thomas¹, (1)Colorado State University, Fort Collins, CO, (2)University of Guelph, Department of Animal Biosciences, Centre for Genetic Improvement of Livestock, Guelph, ON, Canada, (3)Department of Animal Science, University of California, Davis, CA, (4)University of California-Davis, Davis, CA, (5)University of Colorado at Denver, Anschutz Medical Campus, Aurora, CO

Our understanding of the pathology underlying pulmonary hypertension (PH) of cattle in high altitude (> 1,500m) production systems is limited. Chronic PH can result in abnormal cardiac function and heart failure of susceptible cattle. Therefore, RNA-Seq analysis of cardiac tissues was performed to discover differentially expressed genes between hypertensive and normotensive steers (n=7/group). The RNA-Seq analysis revealed 650 differentially expressed genes (P < 0.05). Functional analysis using Ingenuity Pathway Analysis software identified differentially expressed genes related to calcium utilization and availability as important within the gene subset. Quantitative RT-PCR was utilized to validate the expression of 10 candidate genes in cardiac muscle tissues from Angus steers (n=10/treatment group). These genes were selected based on their calculated fold-change differences between the hypertensive and normotensive steers in the RNA-Seq analysis. The selected genes were *ASIC2*, *EDN1*, *FBN1*, *KCNMA1*, *NOX4*, *PLA2G4A*, *RCAN1*, *RGS4*, and *THBS4*. Expression differences (P < 0.0055) existed between hypertensive and normotensive steers for *ASIC2*, *EDN1*, *NOX4*, *PLA2G4A*, *RCAN1*, and *THBS4* in right ventricle samples. Additionally, right papillary muscle exhibited expression differences between hypertensive and normotensive steers for *NOX4*, *PLA2G4A*, *RCAN1*, and *THBS4* (P < 0.0055). These results identify and validate differential expression of genes that may be of interest when evaluating the role of calcium regulation in cardiac tissues pertaining to PH status in Angus steers at high altitude.

PE0370: Cattle

Alteration of mtDNA Copy Number, Mitochondria-Related Gene Expression and Metabolites in Grass-Fed and Grain-Fed Angus Cattle

Ying Bai¹, Jose Carrillo², Yanghua He¹, Yaokun Li¹ and Jiuzhou Song¹, (1)University of Maryland, College Park, MD, (2)CDCB - Council on Dairy Cattle Breeding, Bowie, MD

Grass-fed and grain-fed beef differ in several nutrients. However, the mechanism is still unclear. Different dietary sources have been suggested to have different effects on mitochondrial function and dynamic behavior. In response to changes in energy demand and supply, the organism regulates mitochondrial metabolic status to coordinate ATP production, which would significantly influence whole body metabolism and gene expression. Here, we explored the mtDNA copy number, mitochondria-related gene expression, and metabolic bio-markers in grass-fed and grain-fed Angus cattle. We found that grass-fed group had higher mtDNA copy number than grain-fed group. Comparing tissues, mtDNA copy number was higher in the liver than muscle, rumen and spleen. Then, RNA-sequencing data of the four tissues was used to quantify mtDNA and nuclear DNA-related to mitochondria expression. The results showed analogously lower expression of mtRNA in grass-fed group (relative to grain-fed group) across different tissues. For the differentially expressed mitochondria-related nuclear genes, most of them were up-regulated in muscle of grass-fed group and down-regulated in other three tissues, compared to grain-fed group. We found that *COX6A2*, *POLG2*, *PPIF*, *DCN*, and *NDUFA12*, involved in ATP synthesis, mitochondrial replication, transcription and maintenance, might contribute to the alteration of mtDNA copy number and gene expression in grass-fed and grain-fed steers. Meanwhile, 40 and 23 metabolic bio-markers were identified in blood and muscle of grain-fed group compared to grass-fed group, respectively. Integrated analysis of the altered metabolites and gene expression revealed *SHMT1* and *MDHI*, belonged to the glyoxylate and dicarboxylate metabolism pathway, might regulated glucose and energy metabolism in grass-fed and grain-fed cattle. Taken together, the results provide a basis to further elucidate the adaptive and regulatory modulation of the mitochondrial function in response to different feeding systems in Angus cattle.

PO0371: Cattle

SNP and Haplotype-Based Genome Wide Association Study for SNPs Associated with Tick Tolerance in F2 Nguni x Angus Cattle.

Nelisiwe Mkize¹, Ntanganedzani O. Mapholi², Bekezela Dube¹, Azwihangwisi Maiwashe¹ and Kennedy Dzama³, (1)Agricultural Research Council, Pretoria, South Africa, (2)University of South Africa, Pretoria, South Africa, (3)Stellenbosch University, Stellenbosch, South Africa

The presence of bovine tick is a global problem for cattle industry. The chemical methods in use are not effective and are associated with various disadvantages which threats animal and human welfare. A prospective approach would be to explore genetic variability for bovine tick resistance, since is heritable. Moreover, the availability and inclusion of cattle genetic resistance to ticks in selection programs is a promising strategic method for controlling ecto parasites.

The aim of the present study was to identify SNPs associated with tick resistance in 216 F₂ Nguni x Angus cattle artificial infested with *Amblyomma hebraeum* ticks. Using the SNP and haplotype-based genome wide association approach. The tick count (x) data was not normally distributed, therefore it was transformed using log (x +1), and the transformed data was then tested for normality. Hair samples were used as source of DNA, and the extracted DNA was genotyped using the Illumina BovineSNP150 assay. The obtained genotypes were quality controlled (QC) using Plink software (call rate > 90%, minor allele frequency > 0.01). The SNP based GWAS analysis was implemented using the GenABEL package in R program and the finding were compared to the haplotype based analysis which was conduct on Plink1.9 software.

More data is required to increase the power to validate regions identified in this study. The knowledge and understanding of cattle genetic resistance to ticks will promote effective selection criteria in breeding programs.

PE0372: Cattle

Estimating Genetic Diversity in Vermont Moose Using SNP

Elias Rosenblatt¹, Kimberly M. Davenport², Suraj Bhattarai¹, Veronica Basnayake³, Ted Kalbfleisch⁴, Brenda M Murdoch², Timothy P.L. Smith⁵, James Murdoch¹, Michael P. Heaton⁶ and **Stephanie McKay¹**, (1)University of Vermont, Burlington, VT, (2)Department of Animal and Veterinary Science, University of Idaho, Moscow, ID, (3)GeneSeek, Inc., Lincoln, NE, (4)University of Louisville, Louisville, KY, (5)USDA, ARS, USMARC, Clay Center, NE, (6)USDA, ARS, U.S. Meat Animal Research Center (USMARC), Clay Center, NE

Lack of heterozygosity in moose population may have implications for fitness, which could impact individual responses to parasites and habitat fragmentation. Ultimately, low levels of genetic diversity compromise the population's ability to adapt to changing environmental conditions and reinforce the need for management strategies to incorporate goals for increasing genetic variation. Here, we use a set of single nucleotide polymorphisms (SNPs) that we previously identified in North American Moose, to characterize genetic variability of the moose population in Vermont. An assay consisting of 141 SNPs was used to genotype 179 Vermont moose samples, producing 25,239 total genotypes with a genotype success rate of >90% per animal and a call rate of >90% per loci. The final, curated data set included 136 SNPs from 152 moose for population genetic analyses. Inbreeding and population stratification were estimated and suggest overall low heterozygosity and a high level of identity by state (mean = 0.74) measures for the Vermont moose population. Related animals were identified across northern and central Vermont using pair-wise exclusion, confirming long distance dispersal typical for this species. These results suggest that management decisions should favor strategies to increase diversity, and indicate that the SNP resource provides a useful tool in developing these strategies.

PO0373: Cattle

De novo Haplotype Phased Genome Assembly and Genomic Selection of Buffaloes in India

Ananthasayanam Sudhakar¹, Harish Kothandaraman², Nilesh Nayee³, Sujit Saha³, Dushyant Singh Baghel⁴, Swapnil Gajjar³, Benjamin D. Rosen⁵, Curtis P Van Tassell⁶ and Michael C. Schatz⁷, (1)National Dairy Development Board, Anand, India, (2)Nucleome Informatics Private Limited, Hyderabad, India, (3)National Dairy Development Board, India, (4)Nucleome Informatics Pvt Ltd, Hyderabad, Telangana, India, (5)Animal Genomics and Improvement Laboratory, USDA, (6)Animal Genomics and Improvement Laboratory, USDA, (7)Johns Hopkins University, Baltimore, MD

Here, we report developing a haplotype phased highly contiguous near complete genome assembly of *Bubalus bubalis* by separating haplotypes prior to assembly using a father-mother-offspring trio to accurately and completely reconstruct parental haplotypes.

Blood sample from a typical true to the breed Murrah heifer with known pedigree and normal Karyotype was collected and high molecular weight DNA was isolated. Sample was also collected from both the parents. Parental DNA samples were sequenced on the Illumina platform to generate a total of 274 Gb paired-end data. The progeny DNA sample was sequenced using PacBio long reads (217.3 GB) and 10x Genomics linked reads (226.38GB) along with 802 Gb of optical mapping data and 40X Illumina paired end data

Initially the data was partitioned at a kmer of 21 and then subsequently at 18 considering a minimum kmer coverage of 10 using Meryl. The PacBio Long-read segregation, for each parental haplotype, was over 99.99%, with less than 140Mb of raw reads being present in the non-classified set, confirming the trio. Trio binning based FALCON assembly of each haplotype was scaffolded with 10x Genomics reads and super-scaffolded with BioNano Maps to build reference quality assembly of parental haplotypes. Paternal haplotype assembly was of 2.63Gb with 59 scaffolds whereas maternal haplotype assembly was 2.64Gb in size with 64 scaffolds. N50 observed for both paternal and maternal haplotype was 81.98Mb and 83.23Mb, respectively. The assemblies were submitted to National Centre for Biotechnology Information (NCBI) database - Sire Haplotype accession number VDCB00000000 and Dam Haplotype accession number VDCC00000000. BUSCO single copy conserved core gene set coverage with other eukaryotes genomes was > 91.25%, and gVolante-CEGMA completeness was >96.14% for both haplotypes. Finally, RaGOO was used to order and build the chromosomal level assembly with 25 scaffolds

and N50 of 117.48 Mb (sire haplotype) and 118.51 Mb (dam haplotype). The final haplotype resolved Murrah genome assembly achieved > 99% genome coverage against estimated buffalo genome size of 2.66 Gb.

Whole genome sequencing of 296 buffaloes of 9 riverine breeds and 1 swamp buffalo breed was performed to study the variability among Indian buffaloes. More than 14 million variants including 12 million SNPs were identified per breed after QC, against Mediterranean water buffalo genome assembly (GCA_003121395.1).

Using these variants a custom designed microarray panel BUFFCHIP was developed for genotyping of Indian buffaloes with 60K SNPs.

PE0374: Goats

Footprints of Convergent Adaptation to High-Altitude in Ethiopian and Nepalese Small Ruminant Populations

Kwan-Suk Kim¹, Neena Amatya Gorkhali², Hailu Dadi³ and Zewdu Edea Bedada¹, (1)Chungbuk National University, Cheongju, South Korea, (2)Animal Breeding Division-NARC, Kathmandu, Nepal, (3)Addis Ababa Science and Technology University, Addis Ababa, Ethiopia

Humans have migrated to and inhabited extreme high-altitude environments for several thousand years. Globally, the Tibetan(Himalayan), Andean and Ethiopian highland inhabitants have been widely attracted scholars and served as the classic model to study human populations adaptation to high-altitudes. Follow their domestication, sheep and goats have moved with their custodians to new ecological niches including extremely high-altitude environments. Nepalese and Ethiopian indigenous sheep and goat breeds have evolved under high-altitude environments (>2500 m) for thousands of years; and may have been naturally selected to adapt to hypoxia, low temperatures, and high-altitude radiation. In this study, in an attempt to identify the signatures of convergent adaptation in Ethiopian and Nepalese sheep and goat populations to high-altitude environments, we sampled 4 goat populations genotyped with the Illumina Caprine 50K SNP BeadChip and 5 sheep populations genotyped with the Illumina Ovine 600K chips. By comparing the genomes of sheep and goat breeds evolved under high- and low-altitude environments, we identified signals of selection in regions harboring genes playing an important role in hypoxia adaptation and cellular response to stress. The list of the candidate genes included *PMS1*, *IGF1*, *CAMK2D*, *XRCC4*, *PAPPA*, *BARCA2*, *ATXN2*, and *SCFD1*. Intriguingly, the candidate genes *ATXN2* and *CAMK2D* have previously been identified as under positive selection in human populations inhabiting Ethiopian and Tibetan high-altitudes, respectively. The results suggest that Ethiopian and Nepalese indigenous sheep and goat breeds may have adapted similarly to their respective challenging hypoxic environments.

PO0375: Goats

Genetic Signatures of High-Altitude Adaptation in Tibetan Goats

Caihong Wei, Institute of Animal Sciences, Chinese Academy of Agricultural Sciences, Beijing, China

In the process of domestication, goats have undergone long-term artificial and natural selection, leading to differences among goat breeds in production performance, stress resistance, environmental adaptability and other aspects and leaving different selection traces on the genome. However, the genetic components underlying high-altitude adaptation remain largely unknown. Here, based on previous studies, we selected one highland goat breed from Tibetan plateau and three lowland breeds, which had similar genetic backgrounds with Tibetan goat, around the Tibetan plateau. Then, we genotyped these four breeds using the Illumina Ovine 50K SNP Chip. We found some regions harbouring local positive selection in Tibetan goat by Di statistic, and identified many candidate genes were revealed to be potentially adaptation high-altitude environment. By KEGG enrichment analysis, these candidate genes were enriched in hypoxia, cardiovascular system, energy metabolism and melanogenesis, such as VEGF signalling pathway, EGFR tyrosine kinase inhibitor resistance, and melanogenesis. And ClueGo functional analysis reveals that FGF2, EGFR, AKT1, KDR, PTEN, MITF genes may play some important roles in Tibetan Goat. These findings provide theoretical knowledge for the study of high-altitude adaptability and offer guidance for the molecular breeding of Tibetan goats.

PE0376: Goats

Identification of Selection Signatures in 20 Diverse Goat Breeds

Jan Henkel, Institute of Genetics, Vetsuisse Faculty, University of Bern, Bern, Switzerland

Domestication and human selection have formed diverse goat breeds with characteristic phenotypes. This process correlated with the fixation of causative genetic variants controlling breed-specific traits within regions of reduced genetic diversity, so called selection signatures or selective sweeps. We performed a screen for selection signatures in 20 genetically diverse modern goat breeds and Bezoar goats. We pooled DNA of 12 animals per breed and sequenced the obtained pools to ~30x coverage. The sequence reads were mapped and single nucleotide variants were called. Using sliding windows of 150 kb, we calculated heterozygosity scores and weighted population pairwise F_{ST} values. We identified 5,220 windows with significantly reduced heterozygosity (0.8% of all windows) and 1,474 windows with significant F_{ST} (0.2% of all windows). Adjacent or overlapping windows were further merged. This resulted in 2,239 selection signatures or 1.1% of the total genomic length for the reduced heterozygosity and 847 signatures or 0.4% of the total genomic length for the F_{ST} analysis. We are currently investigating the identified selection signatures for candidate causative variants and traits. So far, we identified six candidate causative variants for breed-specific coat color phenotypes. Interestingly, all six variants are large structural variants. We furthermore identified a candidate causative variant for large body size. An update on the ongoing efforts to identify causative variants for traits under selection will be given at the conference.

PO0377: Goats

Integrated Analyses of Chinese Zhongwei Goat Hair Follicle Development Based on Genome-Wide DNA Methylation and Transcriptomic Profile

Qianjun Zhao, institute of animal sciences, Chinese academy of agricultural science, Beijing, China and PAG Authors' Group

As a rare and precious fur goat breed, Chinese zhongwei goat has a reputation for its pelts worldwide, which are obtained at about 35 days of age. Nevertheless, these natural and exquisite patterns of fleeces are disappearing within two months, causing decreasing economic value with the curly wool straightening. Significantly, the exact reasons of this dynamic transformation are still elusive. DNA methylation plays important roles in mammalian cellular processes and is essential for the initiation of hair follicle (HF) development. Here, we sought to investigate the effects of genome-wide DNA methylation by combining expression profiles of the underlying curly fleece dynamics. In this study we sampled shoulder skin tissues from the same Chinese zhongwei goats in their 45 days and 108 days after birth, to get the DNA methylation landscape and transcriptional expression by conducting whole genome bisulfite sequencing (WGBS) and RNA sequencing (RNA-seq). Between the two developmental stages, 1250 of 3379 differentially methylated regions (DMRs) were annotated in differentially expressed genes (DMGs), and these regions were mainly related to intercellular communication and the cytoskeleton. Integrated analysis of the methylome and transcriptome data led to the identification of 14 overlapping genes that encode crucial factors for wool fiber development through epigenetic mechanisms. Integrated analysis (WGBS and RNA-seq) was performed using differentially methylated regions (DMRs) and differentially expressed mRNAs, and a novel gene was selected to perform the following biological function validation on human hair follicle inner root sheath (HIRS) cells. Furthermore, our functional study highlights a possible role for a special gene in HF cell growth, which may be a predictable biomarker for fur goat selection.

PE0378: Goats

Genome-Wide Insights on Signatures of Positive Selection Reveals Candidate Genomic Regions for Trypanotolerance in Ethiopian and Cameroon Goat Populations

Getinet Mekuriaw Tarekegn^{1,2}, Kouam Simo Jaures^{3,4}, Patrick Wouobeng^{3,5}, Joram M. Mwacharo⁶, Appolinaire Djikeng⁷, Raphael Mrode⁸, Bin Liu^{9,10}, Wenguang Zhang¹⁰, Mwai Okeyo⁸, Dessie Tadelle¹¹, Kassahun Tesfaye^{12,13}, Josephine Birungi⁴, Dirk-Jan de Koning¹, Erling Strandberg¹, Zewdu Edea Bedada¹⁴ and Félix Meutchieye^{3,4}, (1)Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences, Uppsala, Sweden, (2)Department of Animal Production and Technology, Bahir Dar University, Bahir Dar, Ethiopia, (3)University of Dschang, Dschang, Cameroon, (4)Biosciences

eastern and central Africa-International Livestock Research Institute Hub, Nairobi, Kenya, (5) Biosciences eastern and central Africa-International Livestock Research Institute Hub, Nairobi, Cameroon, (6) International Centre for Agricultural Research in Dry Areas, Addis Ababa, Ethiopia, (7) Centre for Tropical Livestock Genetics and Health, The University of Edinburgh, Edinburgh, United Kingdom, (8) International Livestock Research Institute, Nairobi, Kenya, (9) Nei Mongol BioNew Technology Co.Ltd, Hohhot, China, (10) College of Animal Sciences, Inner Mongolia Agricultural University, Hohhot, China, (11) International Livestock Research Institute, Addis Ababa, Ethiopia, (12) Department of Microbial, Cellular and Molecular Biology, Addis Ababa University, Addis Ababa, Ethiopia, (13) Ethiopian Biotechnology Institute, Addis Ababa, Ethiopia, (14) Chungbuk National University, Cheongju, South Korea

Trypanosomiasis is the main livestock health challenge in tsetse-fly infested areas in Africa. 78% of goats in the continent are found in a trypanosomiasis risk zone. This gives a reason for exploring candidate genomic regions related to trypanosomiasis. We genotyped 367 animals from Ethiopia (Keffa and Gumez (tsetse-fly area goats), and Arsi-Bale (Alpine type)), Cameroon (North-west Highland-(NWH)) and China (Cashmere goat) using 50K SNP chip. We used the F_{ST} approach to uncover candidate genes related to trypanosomiasis. The comparison revealed many genes based on significantly differentiated outlier SNPs (at 1%) that are involved in a wide range of biological processes and pathways. To mention some, 19 genes in GTPase activator activity pathway that include ARAP2, ARHGAP15, ARHGAP20, ARHGAP24, ARHGAP26, ARHGAP32 genes and 3 genes in Rac-GTPase binding. Interestingly, we observed highest differentiations ($F_{ST}=0.403-0.478$) among Cashmere against Cameroon and Ethiopian goats but lowest differentiations ($F_{ST}=0.005-0.022$) for Cameroon against Ethiopian goats on ARHGAP15 ([rs268276992](#)). ARHGAP15 is one of the candidate genes for trypanosomiasis in GTPase activator activity molecular function. RAC and VAV gene families in T-cell receptor pathway (an immune response pathway) were highly differentiated between Cashmere with Cameroon and Ethiopian goats ($F_{ST}=0.39-0.76$), but showed weak differentiation between Ethiopian and Cameroon goats ($F_{ST}=0.01-0.10$). ARHGAP15 is an antagonist and negative regulator for RAC, where the latter shows high expression in trypanosomiasis sensitive populations, whereas VAV is an agonist for RAC. In conclusion, the results provide a basis for further research on the genomic characteristics of NWH, Keffa and Gumez goats for trypanosomiasis resistance.

PO0379: Sheep

International Sheep Genomics Consortium Update

Shannon M. Clarke, AgResearch Ltd. Invermay Agricultural Centre, Mosgiel, New Zealand

The long-term goals of the International Sheep Genomics Consortium (ISGC) to develop underpinning resources for the sheep research community has resulted in continued improvement of the sheep genome assembly and development of low, medium and high-density Illumina SNP chips. The ISGC members have continued to make available whole genome sequence data to the community that has been captured via the Sheep Genomes Database, an initiative of the ISGC that extends the consortiums recent achievements. SheepGenomesDB is an electronic warehouse containing sequence variants called from the expanding collection of sheep genomes. Through the application of a single harmonised pipeline for read QC, mapping, variant detection and annotation, SheepGenomesDB makes available variant collections derived in a standardised manner. Run 2 has seen ~1000 animals analysed with variant collections positioned on the OAR V3.1. The consortium is now in the process of Run 3 that will include an additional ~300 animals utilising Rambouillet v1 genome assembly with the aim of providing users with tools to obtain variants defined by chromosomal location, SNP annotation results or via animals and breeds of interest. An update of the ISGC's activities will be presented.

PO0381: Sheep

Identifying Novel Variants in Economically Important Genes through the Flock54 Low-Density Sheep Genotype Panel

Gabrielle M. Becker¹, Kimberly M. Davenport¹, Abby L. Davis¹, Mingrui Duan¹, Lesa Eidman², Christopher S. Schauer³, Whit C. Stewart⁴ and Brenda M. Murdoch¹, (1) Department of Animal and Veterinary Science, University of Idaho, Moscow, ID, (2) Superior Farms, Sacramento, CA,

(3)Department of Animal Science and School of Natural Resource Sciences, North Dakota State University, Hettinger, ND, (4)Department of Animal Science, University of Wyoming, Laramie, WY

Flock54 is a low-density genotype panel designed to provide sheep producers with a cost-effective option for disease carrier and predisposition testing, in addition to testing for variants associated with economically important traits. The panel was designed with causative variants curated from public archives and includes markers for scrapie, spider lamb, rickets, and lentivirus susceptibility, among others. Markers for production traits include fecundity variants associated with *BMP15*, *BMPR1B*, and *GDF9* genes, as well as variants for Callipyge and yellow fat. Amplicon sequences from the Flock54 panel were analysed to identify novel variants in 379 genotyped sheep. Over 1,800 novel variants were identified from markers distributed evenly across the 26 autosomes and the X chromosome. Eleven rickets-affected rams from Wyoming and North Dakota were genotyped along with their cohort (n=158), as Flock54 includes the *DMP1* variant that was identified as causative for inherited rickets in New Zealand Corriedale sheep. There was no variation found in the marker itself, but analysis of the amplicon sequence identified a novel A>AG insertion in codon 125 of *DMP1* which was found to be significantly associated with Wyoming and North Dakota rickets-affected sheep ($P = 0.041$). The described insertion is expected to result in a premature stop codon at codon 137 of *DMP1*. This study describes analysis of novel variants identified in economically important genes of US sheep, and demonstrates the utility of the Flock54 genotype panel for novel variant discovery.

PE0382: Sheep

Chasing Colors: Identifying the Genetic Variants Responsible for Coat Color Variation in Sheep

Christian J Posbergh¹, Elizabeth A. Staiger^{1,2} and Heather J Huson¹, (1)Cornell University, Ithaca, NY, (2)Auburn University, Auburn, AL

White wool is the dominant product in the commercial wool market and commands a higher price than non-white wool. However, non-white wool can bring significantly higher prices than white wool in the hand spinning wool market, particularly in the Northeastern United States. Several studies have examined the phenotypic variation and genetic inheritance patterns at specific loci including the Agouti, Extension, and Brown loci in sheep. However, there has been little work to determine the molecular cause of these different phenotypes. The objective of this study was to identify variants associated with various coat color variation observed within and across sheep breeds, such as dilution, red head and legs, and white spotting. Using Illumina 150 base pair paired-end reads we generated whole genome sequences (approximately 20x coverage) from 18 sheep across the Romeldale, Romney, Jacob, and California Red breeds that were selected to represent a variety of coat color patterns. Reads were aligned to the Oar_v4.0 genome assembly using the Burrows-Wheeler aligner, and variants called following Genome Analysis Toolkit's "Best Practices" workflow. To date, only genes known to be associated with color variants in other species have been investigated. We have identified a variant responsible for the lilac dilution in Jacob sheep within *MLPH*, and a variant within *TYRP1* which is associated with the red hair color observed within the California Red and Tunis sheep. Work is still ongoing to identify variants around the *ASIP* region due to the wide variety in phenotypes attributed to the *ASIP* locus.

PO0383: Sheep

Whole-Genome Sequencing and Phylogenomic Analysis of Wild Thinhorn Sheep (*Ovis dalli*)

Sarah Helen Dias dos Santos¹, Si Lok², Timothy N. H. Lau² and David W. Coltman¹, (1)University of Alberta, Edmonton, AB, Canada, (2)The Centre for Applied Genomics, Peter Gilgan Centre for Research and Learning, The Hospital for Sick Children, Toronto, ON, Canada

Thinhorn sheep, the pure white Dall (*Ovis dalli dalli*) and the dark pelage Stone subspecies (*Ovis dalli stonei*), are Canadian icons. The two subspecies are believed to have arisen through vicariance at the last ice age. Due to admixture following the retreat of the ice sheets, their complex evolutionary relationship is not fully understood. To provide a genomic underpinning to study their relationship, our group is embarking on whole genome sequencing and *de novo* assembly of wild thinhorn sheep. Here, we present the first draft genome assembly of the Stone sheep generated by short-read sequencing combined with microfluidic partitioning on the 10X Genomics platform. The resulting highly contiguous assembly has scaffold N50 length of over 8 Mbp and total reconstruction of 2.6 Gbp, representing 87% of the genome estimated by K-mer analysis. Ninety-two percent of BUSCO mammalian gene set

is represented in its entirety. We have also assembled the complete mitochondrial genomes for Dall and Stone sheep and have performed phylogenetic analysis with other *Ovis* species. Genome assembly of the Dall sheep along with improved draft assemblies of the Stone and intermediate “Fannin” sheep are in progress. When completed, we first aim to identify the genomic determinants linking pelage colour and other notable traits, with the longer-term goal to reconstruct the comprehensive evolutionary and phylogenomic history of Canadian wild sheep and their admixture in a cultural, geographical and environmental perspective.

PE0384: Sheep

Hitting the Mark: Characterizing Four Histone Modifications in Ovine Liver and Spleen

Kimberly M. Davenport¹, Alisha T. Massa², Michelle R. Mousel³, Maria K. Herndon⁴, Stephen N. White³, Suraj Bhattarai⁵, Stephanie McKay⁵, Timothy P.L. Smith⁶, Jacky K. Carnahan⁷, Noelle Cockett⁸, Brenda M. Murdoch¹ and the Ovine FAANG Project, (1)Department of Animal and Veterinary Science, University of Idaho, Moscow, ID, (2)Washington State University, Pullman, WA, (3)USDA, ARS, Animal Disease Research Unit, Pullman, WA, (4)Department of Veterinary Microbiology and Pathology, Washington State University, Pullman, WA, (5)University of Vermont, Burlington, VT, (6)USDA, ARS, USMARC, Clay Center, NE, (7)USDA, ARS, U.S. Meat Animal Research Center (USMARC), Clay Center, NE, (8)Utah State University, Logan, UT

Although the ENCODE project has defined regulatory elements in humans, very little is known in sheep. The functional annotation of the sheep genome, including the location of gene regulatory elements, is essential for understanding the potential mechanisms that may influence economically important traits. The objective of this study is to identify the locations of gene regulatory elements in sheep by characterizing histone modifications in two tissues, liver and spleen. Tissue samples were collected from two female and two castrated male sheep and flash frozen in liquid nitrogen. Chromatin immunoprecipitation (ChIP) was conducted for four histone marks; H3K4me3, H3K27ac, H3K4me1, and H3K27me3 known to be related to active or repressed chromatin states of gene regulatory elements. ChIP sequencing libraries were prepared and sequenced to >60 million reads each. Quality control parameters were performed with FastQC and Trim Galore, and high-quality reads were mapped to Oar_rambouillet_v1.0 with Bowtie2. Peaks were called for narrow marks using MACS2 and broad marks using SICER with false discovery rates of 0.05 and 0.01, respectively. Similarity between animals were examined using a Spearman correlation. Peaks were compared between tissues and sexes to identify states of conservation and differences. Chromatin states were characterized by implementing a Hidden Markov Model with 15 states in ChromHMM. The study identified genomic positions enriched for these four histone marks in sheep, establishing the likely boundaries of regulatory elements involved with gene regulation in these two important tissues. These results will aid in future identification of regulatory mechanisms that influence economically important traits.

PO0385: Sheep

Comparative Genomics of Sheep Tas2r Genes to Cattle, Goat, Human, Dog and Mice

Dillan Henslee, University of Idaho, Hagerman, ID, Brenda M. Murdoch, Department of Animal and Veterinary Science, University of Idaho, Moscow, ID, Joel Yelich, University of Idaho, Carmen, ID, Bret Taylor, USDA-ARS Sheep Experiment Station, Dubois, ID and Melinda J. Ellison, University of Idaho, Moscow, ID

Type Two Taste Receptors (Tas2r) are the only taste receptors that distinguish bitter-tasting compounds. Human Tas2r genes have been extensively studied and are associated with diet preference, health, substance abuse, and dietary habits. Sheep are an important livestock species known for grazing vast rangelands with variable ecology and plant communities. However, the limited work in Tas2r gene repertoires in the reference genomes of grazing animals creates a challenge for understanding how these genes influence diet selection preferences. Tas2r genes cluster on two separate regions of the genome. In cluster 2 of the sheep (OAR_rambouillet_1.0), goat (ARS1), and cattle (ARS-UCD1.2) reference genomes, there are six, nine, and two genes are not annotated, respectively. A nucleotide similarity comparison of the whole Tas2r repertoires for these three grazing species suggests goat and cattle are similar to sheep ($\geq 95.5\%$ and $\geq 91.9\%$ similarity, respectively). Comparative genomic strategies were used to cross-reference sheep Tas2r genes with five other species for the proposed annotation of cluster 2 including 42, 67, 67B, 31, 12, 10B, 10, 9, 8, and 7. Genes 42, 9, 8, and 7 have previously been annotated in sheep, where as

67B and 10B are likely not found in human, dog, or mice and may be secluded to ruminants. This technique was also used to improve annotation of goat and cattle repertoires. Understanding how Tas2r genes may influence diet selection in grazing ruminant species could provide more insight into management of western rangelands through sheep grazing strategies.

PE0386: Sheep

Effect of Heat Stress and Beta Adrenergic Agonist Supplementation on the Adipose Transcriptome of Lambs

Rachel R. Reith, Erin M. Duffy, Rebecca M. Swanson, Anna M. Fuller, Ty B. Schmidt, Dustin T. Yates and Jessica L. Petersen, University of Nebraska-Lincoln, Lincoln, NE

Beta adrenergic agonists (BAA), supplemented to finishing livestock, improve feed efficiency and decrease carcass fat. Conversely, environmental heat stress reduces animal performance. How these treatments may interact, however, and how BAA supplementation itself impacts genomic activity is not fully understood. The purpose of this study was to use RNA-sequencing to gain a better understanding of how BAA supplementation and heat stress alter gene expression of subcutaneous adipose. Lambs were used as a model of ruminant livestock to examine the impact of ractopamine HCl, a beta one adrenergic agonist, and heat stress on adipose tissue gene expression. The lambs were assigned to one of two temperatures (thermal neutral or heat stress) and supplement (no supplement or ractopamine HCl 60mg/head/day) treatments, for a total of 4 groups of 6 lambs each. Daily measurements of intake, respiration rate, and rectal temperature were recorded, and blood was collected every third day throughout the trial. After 30 days, the lambs were slaughtered and subcutaneous fat was flash-frozen. RNA was isolated from the adipose and poly-A selected libraries were sequenced using 150bp, paired-end reads to a targeted depth of 20 million reads per sample. Resulting data were trimmed for quality, and transcript counts as annotated in Oar_rambouillet_v1.0 were quantified using STAR. Differential expression analyses conducted in limma-voom found no significant differences between treatment groups in gene expression. The most highly expressed transcripts, however, allow for hypothesis building for future studies.

PO0387: Sheep

Association of *TMEM8B* with Mature Weight in Sheep

Mehmet Ulas Cinar^{1,2}, Michelle R. Mousel^{3,4}, Maria K. Herndon¹, J. Bret Taylor⁵ and **Stephen N. White^{1,3}**, (1)Department of Veterinary Microbiology and Pathology, Washington State University, Pullman, WA, (2)Faculty of Agriculture, Department of Animal Science, Erciyes University, Kayseri, Turkey, (3)USDA, ARS, Animal Disease Research Unit, Pullman, WA, (4)School of Global Animal Health, Washington State University, Pullman, WA, (5)USDA, ARS, Range Sheep Production Efficiency Research, US Sheep Experiment Station, Dubois, ID

Multiple signature-of-selection studies have identified single nucleotide polymorphism (SNP) *rs426272889* as having been under extreme selection pressure in the recent evolutionary history of domestic sheep. However, no study had provided and explored phenotypic data suggesting a functional reason associated with this extreme selection. We tested *rs426272889* for association with a panel of production traits in 779 U.S. Rambouillet, Targhee, Polypay, and Suffolk sheep. We identified association with mature weight at ages 3 and 4 years ($P < 0.05$). This confirmed that sheep body size may be a primary reason for the extreme historical selection in this genomic region. The region surrounding SNP *rs426272889* has had updated genomic annotation, and it is now identified in *TMEM8B*. *Transmembrane protein 8B* has little functional annotation in any mammalian species beyond a role inhibition of cancer cell proliferation. To our knowledge, this is the first study to link *TMEM8B* to body size under normal extensive production conditions. Additional work is required to identify the underlying functional variant or variants affecting body size. Once functional variants are identified, they may be used for selective breeding to improve sheep production.

PE0388: Sheep

A Genome-Wide Association Study (GWAS) of the Multi-Horned Trait in Sheep Based on Whole-Genome Re-Sequencing

Xiaohong He, Institute of Animal Science, CAAS, Beijing, China, Lin Jiang, Institute of Animal Science, CAAS, Beijing, China and Yuehui Ma, Institute of Animal Science, Chinese Academy of Agricultural Sciences (CAAS), Beijing, China

Multi-horned sheep is an ideal animal model for illuminating the genetic basis of horn development. After observing the multi-horned phenotype within Tibetan sheep at an altitude of 5200 meters above sea level (m.a.s.l.), we have found four sheep breeds with multi-horned phenotype among native sheep breeds in China. In the study, a genome-wide association study (GWAS) were performed to identify the genetic region of multi-horned trait in 16 sheep breeds of 394 individuals. The samples were sequenced on Illumina HiSeq XTen/2500 instruments. Our genome data set includes 4 Chinese multi-horned breeds, 8 Chinese indigenous breeds and 3 breeds from Africa and Europe and one wild breed. The 4 multi-horned sheep breeds in this study include Altay sheep, Mongolian sheep, Sishui Fur sheep and Tibetan sheep which are ancient sheep breeds in China. The median genome coverage achieved across the full data set was $\sim 7.5\times$. High-quality trimmed read pairs were aligned against the reference sheep genome assembly Oar_v4.0 genome using BWA (version: 0.7.12) with default parameters. Strict read alignment and genotyping calling procedures allowed us to identify a total of 17,075,402 single nucleotide polymorphisms (SNPs). A case-control design GWAS was performed to investigate the genetic basis of the multi-horned phenotype. The compressed mixed linear model program GAPIT V2.12 was used for the association analysis. A single strong association signal was observed on chromosome 2. 75 percent of all 527 significant SNPs were located on chromosome 2. The top 10 associated SNPs were located in the 132.59–133.57 Mb and the P value of the most significant SNP was 8.19×10^{-29} . The results narrow the genetic region of the DNA chip analysis in the previous studies. We also performed the CNV analysis, but there were no CNVs found in candidate region on chromosome 2.

PO0389: Sheep

Cytokine/Chemokine and microRNA Profiles of Variable Stress Responding Sheep Under Bacterial Endotoxin Challenge

Ankita Sharma¹, Danielle Naylor¹, Tianna Sullivan¹, Ziwei Li², Kristen Lamers¹, Anegla Canovas³, Bonnie A. Mallard⁴ and Niel A Karrow¹, (1)Centre for Genetic Improvement of Livestock, Department of Animal Biosciences, University of Guelph, Guelph, ON, Canada, (2)AHL University of Guelph, Guelph, ON, Canada, (3)Department of Animal Biosciences, University of Guelph, Guelph, ON, Canada, (4)Department of Pathobiology, Ontario Veterinary College, University of Guelph, Guelph, ON, Canada

Antimicrobial (AM) usage to combat biological stressors has contributed to the development of AM-resistant pathogens. Genetic selection for enhanced stress resilience could be a possible strategy to maintain animal health and reduce AM usage. Variation and heritability ($h^2 \approx 0.3$) of the stress response in sheep to intravenous immune challenged with lipopolysaccharide (LPS) endotoxin (*E. coli* 0111:B4) has been reported previously, and variation was associated with the immune response. The present study aimed to characterize circulating cytokine/chemokine and microRNA profiles of high (**HSR**, cortisol [336.2 ± 27.9 nmol/L]), middle (**MSR**, cortisol [147.3 ± 9.5 nmol/L]) and low (**LSR**, cortisol [32.1 ± 10.4]) stress responders in LPS-challenged sheep ($n=112$, 80 days old) along with rectal temperature. Serum samples were collected at 0 h and 2, 4, 6 h post-LPS challenge to profile microRNAs (miR-130, miR-200b, miR-31, miR-29, miR-223, miR-1246, miR-145). Further, 15 cytokines/chemokines were measured at the peak cortisol time (4 h) and compared with pre-stress levels. Variable stress-responders demonstrated distinct responses for IL-6, IFN- γ , IL-10, CCL2, and CXCL10 ($p < 0.05$). Rectal temperatures were also distinct ($p < 0.01$), with the HSR having the strongest fever response. The expression of immune-related miR-145, miR-31, miR-223, miR-1246, miR-200b and miR-130 was higher in the LSR ($p < 0.05$), whereas, miR-29b was higher in the HSR. Variable cytokine/chemokine and miRNA profiles among stress responders further supports an association between stress responsiveness and immune health, and this association warrants further investigation if selection for enhanced stress resilience is pursued.

PE0390: Sheep

Effect of the Booroola Mutation at BMPR1B on Range Lamb Production of Finewool Sheep

Jacob W. Thorne, University of Idaho, San Angelo, TX, David G. Riley, Department of Animal Science, Texas A&M University, and Texas A&M AgriLife Research, College Station, TX, College Station, TX, Reid Redden, Texas A&M AgriLife Extension, San Angelo, TX, Brenda M. Murdoch, Department of

Animal and Veterinary Science, University of Idaho, Moscow, ID and Daniel F. Waldron, Texas A&M AgriLife Research, San Angelo, TX

Number of lambs weaned can significantly impact the economic viability of a commercial sheep operation. Texas has the largest sheep population in the United States, yet generally has a 30 percent lower lamb crop than the national average. The Booroola mutation (*FecB*) at gene BMPR1B has been shown to increase ovulation rate and litter size in multiple breeds of sheep, but has not been tested in an extensively managed flock under range conditions common to the Texas sheep industry. Over a 4-year period, 316 grade Rambouillet ewes, 154 of which were heterozygous for *FecB*, were managed as one flock for ten months and only separated, by genotype, for lambing over the remaining two months in similar pastures. Ewes were expected to give birth and rear their lambs unassisted with continual access to native pasture and limited supplemental feeding. Litter size (open, single, multiple) was detected via ultrasonography, and number of lambs weaned and weaning weights of the lambs were recorded. Post weaning, lambs were fed a typical feedlot ration ad libitum and re-weighed at 60 d. Ewes heterozygous for the *FecB* mutation carried multiple offspring more often at mid-gestation (58.3% vs. 41.7%, $p < 0.001$) and tended to wean more lambs (1.28 vs. 1.09, $p < 0.051$), but litter weights at weaning did not differ between the two groups as lambs produced by wild-type ewes were significantly heavier at weaning (26.36 kg vs 22.29 kg, $P < 0.05$). The weight advantage of lambs produced by wild-type ewes remained significant after 60 d post weaning (42.84 kg vs. 38.06 kg, $p < 0.001$). In summary, the influence of the Booroola mutation did not increase the overall flock productivity under Texas range conditions.

PO0391: Sheep

Whole Genome Neutral and Adaptive Diversity in Traditional Sheep Breeds: Case of Morocco

Badr Benjelloun^{1,2}, Abdessamad Ouhrouch^{3,4}, Simon Boitard⁵, Abdelmajid Haddioui³, Frédéric Boyer⁶, Bertrand Servin⁷, François Pompanon⁸ and Anne Blondeau Da Silva⁹, (1)LECA, UGA-CNRS-USMB, France, (2)INRA-Morocco, Beni Mellal, Morocco, (3)USMS Beni Mellal, Morocco, (4)INRA-Morocco, Morocco, (5)GenPhySE, INRA-INPT-ENVT-Université de Toulouse, France, (6)LECA/CNRS, Grenoble, France, (7)GenPhySE, INRAE, ENVT, ENSAT, Toulouse, France, (8)LECA - Université Grenoble Alpes, Grenoble, France, (9)GAMAA Univ. Limoges, INRA, France

Genomic selection signatures are very useful for understanding how environmental and anthropic constraints impact genome variation and the distribution of domestic species. Sheep farming plays a key role and is one of the main sources of milk and meat in the world. Unlike cosmopolitan breeds, traditional populations represent interesting hotspot of diversity and the understanding of their distribution would help facing the impact of environmental changes. Here, we characterized neutral genome diversity and demographic history as well as intra and inter-population selection signatures in the main sheep breeds reared in Morocco using their entire genomes. The complete genome data from 87 individuals representing five predominant local sheep breeds in Morocco were used to infer demographic history, which has made it possible to estimate the evolution of the effective population size over time. Two methods were used to investigate selection signatures: one to detect putative regions under selection within each of these breeds and the second to detect selection signatures that differentiate the breeds one from the other. We identified several hundreds of regions/genes under selection from the studied breeds. We highlighted several biological processes involved in local adaptation as well as those linked to zootechnical performances characterizing each breed. Findings of this study increased our understanding on how genetic diversity is distributed in local breeds.

PE0392: Sheep

Considerations for Designing Imputation Studies Using Whole Genome Sequence Data: A Study in a Diverse New Zealand Sheep Population

Andrew S Hess¹, Rudiger Brauning¹, Michael Lee², Melanie k. Hess³ and Shannon M. Clarke⁴, (1)AgResearch Limited, Invermay Agricultural Centre, Mosgiel, New Zealand, (2)University of Otago, Dunedin, New Zealand, (3)AgResearch Limited, Invermay Agricultural Centre New Zealand, Mosgiel, New Zealand, (4)AgResearch Ltd. Invermay Agricultural Centre, Mosgiel, New Zealand

Whole-genome sequencing (WGS) provides an in-depth view of the genetic variation that is present in an individual and population but is cost-prohibitive. Imputation to WGS is an invaluable tool to capture variation in silico in greater detail across the genome while minimizing costs. The ability to impute individuals of interest up to WGS is dependent on how well the animals with WGS data capture the genetic variation present in that population. The International Sheep Genomics Consortium (ISGC) has sequenced ~1000 sheep to explore the global diversity in sheep. Of these animals, ~200 sheep are from New Zealand, which were chosen to represent the genetic diversity in New Zealand, while also requiring that SNP array data (50K or 600K) was available. The goal of this study was to use the SNP array genotype data available on these sheep to assess their potential to impute sheep in New Zealand. We used imputation of 3K SNP array genotypes up to ~42K (the overlap between 50K and 600K SNP arrays) to assess expected accuracy of imputation to WGS and identify key factors influencing imputation accuracy. We also identified other individuals within the New Zealand sheep population that may be of interest to sequence to better capture the diversity in this population. This information can be used to narrow the set of individuals with SNP array genotypes (~190,000 available) to a high-confidence set for GWAS, thus reducing the likelihood of spurious associations due to imputation errors and reducing computational burden.

PO0393: Swine

The EU-H2020 Project GENE-SWitCH (The regulatory GENome of SWine and CHicken: functional annotation during development)

Elisabetta Giuffra, INRA, UMR de Génétique Animale et Biologie Intégrative, Jouy-en-Josas, France and The GENE-SWitCH Consortium

The H2020 project GENE-SWitCH started in July 2019 and will span 4 years. It aims to deliver new underpinning knowledge on the functional genomes of two main monogastric farm species (pig and chicken) and to enable immediate translation to the pig and poultry sectors. The activation status of functional genome sequences varies across time and space, and in response to environmental perturbations. In full coordination and synergy with global effort and ongoing projects of the FAANG community, we will characterize the dynamics (“switches”) of the functional genome from embryo (chicken) and fetus (pig) to adult life by targeting a panel of tissues relevant to sustainable production. New expression QTL data in pigs and existing high-resolution QTL data in chicken will be used for developing innovative genomic predictive models that integrate functional annotations, and these models will be validated in commercial pig and poultry populations. In addition, nutritional epigenetic data will allow evaluation of the influence of maternal diet on the epigenome of the pig fetus and whether such effects persist until post-weaning. These open-shared datasets will conform fully with FAANG standards and add valuable knowledge on genetic and epigenetic variation of functional elements to FAANG. A comprehensive plan of dissemination and outreach activities to a large audience of stakeholders will be implemented. The GENE-SWitCH consortium brings together partners representing pan-European excellence (including the academic institutions which pioneered FAANG) and world-leading animal breeding and biotech industry in a true co-creation effort. Overall, GENE-SWitCH will contribute to the global FAANG effort considerably, demonstrate how functional annotation of genomes can foster the advancement of genomic selection for immediate benefit to the breeding industry, and produce cutting-edge research paving the way to new studies and strategies for sustainable productions. This talk will provide an update of activities carried out in the first 6 months of the project.

PE0394: Swine

Harnessing the Power of the Breed

Thibaut Hourlier, Konstantinos Billis, Jamie Allen, Carlos García Girón, Leanne Haggerty, Osagie Izuogu, Denye Ogeh, Fergal J. Martin, Kevin Howe and Paul Flicek, European Molecular Biology Laboratory - EBI, Cambridge, United Kingdom

Many efforts to dramatically increase the breadth of sequenced eukaryotic genomes are already underway. Concurrently, we are seeing an increase in terms of the depth at which each species is represented, with many alternative assemblies becoming available for key species. Often these assemblies represent different aspects of the biology of a species, for example different breeds, strains, sexes or cell lines.

We are interested in capturing this depth in Ensembl in order to provide robust genomic resources and user-focused services. These help form the basis for new discoveries in the fine-grained aspects of variation within a species such as structural and copy number variants, which can lead to various environmental adaptations and different traits among individuals within the same species.

Since release 98 of Ensembl, users can browse through the genome annotation of twelve breeds of pig and the pig reference genome annotation. The breeds originated from all over the world: Berkshire, Hampshire, Landrace, Large White, Piétrain, Bamei, Jinhua, Meishan, Tibetan Wild Boar, Wuzhishan and a crossbred Duroc/Landrace/Yorkshire. For each of the different breeds, we used the same transcriptomic set comprised of twenty-eight samples of Illumina RNA-seq data (PRJEB19386) and nine samples of PacBio long read data (PRJNA351265) to provide comparable gene sets. We carried supplemental analyses such as pairwise alignments and whole genome alignments, aimed at highlighting the differences and the similarities between the reference assembly and the multiple breed assemblies.

PO0395: Swine

Profiling of Chromatin Accessibility in Developing Pig Muscle to Identify Regulatory Regions

Mazdak Salavati¹, Yennifer Cortés Araya¹, Michelle M. Halstead², Alan L. Archibald¹, John M. Hickey¹, Francisc X. Donadeu¹, Musa A. Hassan¹ and **Emily L. Clark¹**, (1)The Roslin Institute and R(D)SVS, University of Edinburgh, Edinburgh, United Kingdom, (2)Animal Science, University of California, Davis, CA

We have very little information about how the genome is regulated in production pigs. This lack of knowledge undermines efforts to define and predict the effects of genetic variants in pig breeding programmes. In order to address this knowledge gap we need to identify regulatory sequences in the pig genome starting with regions of open chromatin. We have optimised the ‘Improved Protocol for the Assay for Transposase-Accessible Chromatin (Omni-ATAC-seq)’, to profile regions of open chromatin, in flash frozen pig muscle tissue samples. This has allowed us to identify putative regulatory regions in semitendinosus muscle from 24 production pigs, from a developmental time series. The ATAC-Seq data was mapped to Sscrofa 11.1 and Genrich was used for post-alignment peak-calling. We found that >50% of regions of accessible chromatin were within 1kb of known transcription start sites and the width of each region varied according to the developmental time point. In parallel we have measured genome-wide gene expression and allele-specific expression using RNA-Seq to determine the impact of these open chromatin regions on gene expression. The next stage of the study is to leverage the ATAC-Seq data with a very large dataset of genetic variants from production pigs to determine whether any trait-linked variants are located within the open chromatin regions. The dataset we have generated provides a powerful foundation to investigate how the genome is regulated in production pigs and contributes valuable functional annotation information to define and predict the effects of genetic variants in pig breeding programmes.

PE0396: Swine

Genome-Wide Associations for Fatty Acid Composition of Pork Identifies Novel Candidate SNP on Chromosome 14

Dan Nonneman¹, Meghan D. Clancy², Gary A. Rohrer³, Emily Arkfeld⁴, Dustin D. Boler⁵, Anna C. Dilger⁶, Andy King¹, Tommy L. Wheeler¹ and Steven D. Shackelford¹, (1)USDA, ARS, USMARC, Clay Center, NE, (2)Texas Tech University, Lubbock, TX, (3)USDA, ARS, U.S. MEAT ANIMAL RESEARCH CENTER, Clay Center, NE, (4)Triumph Foods LLC, St. Joseph, MO, (5)Topigs Norsvin, Urbana, IL, (6)Department of Animal Sciences, University of Illinois at Urbana-Champaign, Urbana, IL

Fat content and fatty acid composition are attributes that contribute to the nutritional value, sensory properties and quality of pork. Fatty acid composition is dependent upon genetic background, is easily modified through diet and should respond to genetic selection. A genome-wide association study of clear plate (s.c.) fatty acid composition was done in 818 commercial pigs using the GGPHD.v1 beadchip. Thirty-one QTL were identified for fatty acids myristic (14:0), palmitic (16:0), palmitoleic (16:1), stearic (18:0), oleic (18:1(n-9)), arachidic (20:0), SFAs, MUFAs and PUFAs near the genes FADS2 (SSC2), ELOVL5 (SSC7), ELOVL6 (SSC8), SCD and ELOVL3 (SSC14), and ELOVL7 (SSC16) Zero to 3 QTL were identified for individual traits and about 4 to 10 percent of the variance was

explained by markers. Thirty-four potentially functional variants in ELOVL6 on SSC8 (12 SNP), and SCD (10 SNP), HIF1AN (5 SNP), PPRC1 (6 SNP) and ELOVL3 (1 SNP) on SSC14 were genotyped in 718 of the same pigs using Agena MassArray. Twenty-one SNP were located in the 5'-UTR or promoter region and 13 SNP were missense substitutions. The most significant associations were found with SNP in the promoter regions of SCD for 18:0, and HIF1AN for 18:1(n-9), the ratio of 18:1 to 18:0, SFAs and MUFAs. Although SCD and HIF1AN are 158kb apart, low LD was observed between SCD and HIF1AN markers. These results indicate that these markers could be used for selection for increased fatty acid saturation to mitigate pork fat quality issues that affect belly slicing yields and bacon quality.

PO0397: Swine

Genome-Wide Association Study Links Sensory Human Nose Score of Boar Taint to Skatole in Duroc Pigs

Marcos S. Lopes¹, Eli Grindflek², Anette Krogenæs³, Marianne S. Sødning⁴, Gunn Østby³, Inger-Helene Bjørnson Aardal³, Egbert Knol⁵, Hanne Hamland² and **Maren van Son**², (1)Topigs Norsvin, Curitiba, Brazil, (2)Topigs Norsvin, Hamar, Norway, (3)NMBU, Oslo, Norway, (4)Animalia AS, Oslo, Norway, (5)Topigs Norsvin, Beuningen, Netherlands

Surgical castration of male piglets is the most common practice to avoid boar taint (BT) in pork. However, due to animal welfare concerns this procedure is not desirable and other solutions, such as genetic selection, have been sought. In addition, selection against BT has been generally carried out using chemical measurements of androstenone and skatole, which are the major BT compounds. Nevertheless, sensory analysis like human nose score (HNS) of BT measured by trained panelists in a controlled environment has emerged a faster and cheaper alternative to chemical analysis. The aim of this study was to investigate the molecular mechanisms of androstenone, skatole and HNS through genome-wide association analyses in order to identify markers that can assist the selection against BT. Phenotypes and (imputed) Axiom porcine 660K array genotypes were available from a Duroc pig population (N=9740). The results revealed that the significant loci for HNS was also significant for skatole, whereas a QTL region reached genome-wide significance for androstenone on SSC13. For HNS and skatole, the most significant region was located at pig chromosome SSC14 at ~141 Mb. This is close to the gene cytochrome P450 2E1 (CYP2E1), which has previously been found important for levels of skatole in pigs. Another significant QTL region for HNS and skatole was identified on SSC6 at ~6 Mb. This study is the first identifying QTL regions for HNS, and the results show that QTL regions for HNS and skatole are overlapping in Duroc.

PE0398: Swine

Sex Determination in Porcine Using a Single Primer Pair to Determine X- and Y-Chromosome Specific Fragments of GPM6B Homologs

Evan K. Peterson, Miriam Laker and Lee Rickords, Utah State University, Logan, UT

DNA Sex determination of animals is important for embryo transfer and forensic purposes. Such sex determination assays rely on identifying the presence of a Y-chromosome, using PCR amplification. Selecting a single primer pair that will create amplicons of differing sizes from both the X- and Y-chromosome, creates an inherent positive control in every reaction. This reduces the incidence of false negatives without introducing the inefficiencies associated with duplexed PCR reactions. The generation of two distinct amplicons using a single primer pair is accomplished by selecting primers such that the amplified region on one chromosome spans one or more deletion regions.

We have designed a single primer pair for pigs that does this. These primers target the GPM6B gene and its homolog on the Y-chromosome. The GPM6B gene is 213.15 kb long, and located on the reverse strand of the X-chromosome. The transcript contains 10 exons and produces the protein Neuronal membrane glycoprotein M6-b. The protein GPM6B is a four transmembrane protein that is expressed most abundantly in neurons and oligodendrocytes, and is believed to regulate osteoblast function in humans.

The GPM6B Y-chromosome specific homolog contains many deletions, several of which are within the fragment generated by our primers resulting in the production of X- and Y-specific amplicons of differing lengths (244 bp and 201 bp respectively).

PO0399: Swine

Genome-Wide Epistatic Interaction Networks Affecting Feed Efficiency in Duroc and Landrace Pigs

Priyanka Banerjee, Victor AO. Carmelo and **Haja N. Kadarmideen**, Technical University of Denmark, Kgs. Lyngby, Denmark

Interactions among genomic loci have often been overlooked in genome-wide association studies, revealing combinatorial effects of variants on phenotype or disease manifestation. Unexplained genetic variance, interactions amongst causal genes of small effects and biological pathways could be identified using a network or systems biology approach. The main objective of this study was to determine the genome-wide epistatic variants affecting feed efficiency traits [feed conversion ratio (**FCR**) and residual feed intake (**RFI**)] based on weighted interaction SNP hub (**WISH-R**) method. In this study, we detected highly interconnected epistatic SNP modules, pathways, and potential biomarkers for the FCR and RFI in Duroc and Landrace purebreds considering the whole pig population, and separately for low and high feed efficient groups. Highly interacting SNP modules in Duroc (1,247 SNPs) and Landrace (1,215 SNPs) across the population and for low feed efficient (Duroc – 80 SNPs, Landrace – 146 SNPs) and high feed efficient group (Duroc – 198 SNPs, Landrace – 232 SNPs) for FCR and RFI were identified. Gene and pathway analysis identified *ABLI*, *MAP3K4*, *MAP3K5*, *SEMA6A*, *KITLG* and *KAT2B* from chromosomes 1, 2, 5, and 13 underlying ErbB, Ras, Rap1, thyroid hormone, axon guidance pathways in Duroc. *GABBR2*, *GNAI2*, and *PRKCG* genes from chromosomes 1, 3, and 6 pointed towards thyroid hormone, cGMP-PKG and cAMP pathways in Landrace. From Duroc low feed efficient group, *TPK1* gene was found involved with Thiamine metabolism whereas *PARD6G*, *DLG2*, *CRB1* were involved with Hippo signaling pathway in high feed efficient group. *PLOD1* and *SETD7* genes were involved with Lysine degradation in low feed efficient group in Landrace, while high feed efficient group pointed to genes underpinning Valine, Leucine, Isoleucine degradation, and fatty acid elongation. Some SNPs and genes identified are known for their association with feed efficiency, others are novel and potentially provide new avenues for further research. Further validation of epistatic SNPs and genes identified here in a larger cohort would help establish a framework for modeling epistatic variance in future methods of genomic prediction increasing the accuracy of estimated genetic merit for FE and help the pig breeding industry. This is the first study to report epistatic interactions for feed efficiency traits (**FCR** and **RFI**) in pigs providing estimates of genome-wide pair-wise epistasis between SNPs, depicting weighted and scale-free SNP-SNP interaction networks, functional annotations and pathway enrichment of highly connected SNP modules of relevance to RFI and FCR.

PE0400: Swine

Genome-Wide Association Studies for Feed Efficiency with Imputed Genotypes in Divergent Lines of Pigs

Emilie Delpuech, INRA GENPHYSE, CASTANET TOLOSAN, France

Genetic improvement of feed efficiency contributes to the sustainability of animal productions. Two divergent lines on residual feed intake (RFI) are conducted at INRA considering. In each generation, within each line, about 90 males from one sow parity were tested during growth to select the six boars with the lowest (LRFI line) or highest (HRFI line) RFI. Another parity was produced to evaluate the correlated responses to selection on production traits. Genome-wide association studies (GWAS) were performed on these animals for 24 traits: RFI, related traits as average daily gain (ADG), backfat thickness (BFT), and FCR, together with traits related to carcass composition and meat quality. In total, data comprised records from 2,426 response animals with genotypes imputed (with FImpute software V2.2) as the average genotype of the parents for the 570,440 SNP chip.

The GWAS analyses were carried out within line, integrating sequentially individuals of each generation, to trace genomic regions under selection. Results from GWAS revealed 244 significant associations, including 11 regions associated to RFI. Among these QTLs, some confirmed previously published results. Regions detected in line were

different, and *in silico* gene analyses indicated that selection has impacted different metabolic pathways in the LRFI and HRFI lines. These results provide new information on the genetic basis of feed efficiency in pigs.

PO0401: Swine

Allele-Biased Expression and Histone Modification in Fetuses from Reciprocal Crosses of Divergent Pig Breeds

Haibo Liu¹, Buyue Niu¹, Brittney N. Keel², Dan Nonneman³, Timothy P.L. Smith³ and Christopher K. Tuggle¹, (1)Animal Science, Iowa State University, Ames, IA, (2)USDA, ARS, U.S. MEAT ANIMAL RESEARCH CENTER, CLAY CENTER, NE, (3)U.S. MEAT ANIMAL RESEARCH CENTER, CLAY CENTER, NE

Identifying genetic variants affecting gene expression by allele-biased expression (ABE) and epigenetic modification analyses is important for us to understand gene expression regulation. And hybrids of genetically divergent breeds maximize the potential of discovering genetic variants regulating gene expression. In this study, we performed both RNA-seq (4 tissues: brain, liver, muscle and placenta, two developmental stages, Day 30 of gestation (D30G) and Day 70 of gestation (D70G)) and ChIP-seq (liver and muscle, D70G) to identify variants of allele-biased expression (ABE) and allele-biased histone modifications (ABHM), in F1 fetuses from reciprocal crosses of Meishan (MS) and White composite (WC) pigs. Using a robust pipeline for allele-biased analysis, we identified xx SNPs and yy SNPs showing allele-biased expression and histone modifications. Functional annotation of the genes with ABE and ABHM is undergoing. The sets of variants and their associated genes showing ABE and ABHM will provide a foundation for understanding gene regulation and the underlying phenotypic differences between MS and WC.

PE0402: Swine

Whole Genome Analysis of Alentejano Pigs with Contrasting Meat Quality Phenotypes

Ana Usie¹, Hugo Magalhães¹, Célia Leão¹, Daniel Gaspar¹, Brigida Meireles¹, Pedro Barbosa¹, Liliana Cachucho¹, Andre Albuquerque², Rui Charneca², Jose Martins², Eliana Jeronimo¹ and Marcos Ramos¹, (1)Centro de Biotecnologia Agrícola e Agro-Alimentar do Alentejo (CEBAL), Beja, Portugal, (2)Universidade de Évora, Evora, Portugal

The Alentejano is a Mediterranean pig breed, found in southern Portugal, reared under extensive conditions and finished on grass and acorns during the fall and winter months. From these animals a variety of dry-cured meat products of great economic importance are generated. A total of 541 pigs were studied during the 2017 slaughter campaign. Phenotypic records for carcass and meat quality were collected and subsequently analyzed to identify the groups of animals that displayed the most contrasting phenotypes. Two groups comprising 13 animals each were selected, based on pH, water loss, total lipids, total protein, total collagen and pigments content. All samples were re-sequenced to a 23x coverage. The reads were aligned to the pig genome and SNPs and structural variants identified between the two groups of animals. A total of 13,418,254 SNPs were identified, of which 6,851,475 and 6,566,779 were located in the genic and intergenic genomic regions, respectively. The number of SNPs for which at least 25 samples were present comprised 88.7%. The set of genic SNPs included 43,405 exonic non-synonymous SNPs and 60,750 exonic synonymous SNPs. The remaining SNPs were located in introns and ncRNA regions. Interestingly, SNPs with markedly different allele frequencies between the groups were also identified (a total of 230 SNPs with allele frequency differences between the groups of at least 30% or 70%). This study represents the first major characterization of Alentejano pigs at the genome level, and identified a significant number of SNPs potentially associated with meat quality.

PO0403: Swine

Differential Gene Expression across Developmental Stages in Swine Tissues

Suxu Tan, Michigan State University, East Lansing, MI and Wen Huang, Michigan State University, East Lansing, MI

Sus scrofa is one of the most important domestic animals used for meat production and as an excellent model for biomedical research. To better understand the gene expression profiles of the brain, liver, and *longissimus dorsi* muscle, we conducted RNA-Seq to generate an extensive dataset (237 samples) across three development stages (gestational day 40, day 70, and adult) from Yorkshire x Hampshire crossbred individuals. Differentially expressed genes were obtained by pairwise comparison between development stages for each tissue. Functional enrichment analyses were performed using the differential gene list and expressed genes as background gene list. Enriched gene ontology terms and pathways revealed pivotal functions corresponding to tissues, such as neurogenesis in brain, immunity in liver, and muscle system process in muscle. The results showed dynamic expression profiles and suggested the roles of gene expression in swine tissue development.

PE0404: Swine

Identification of Genome-Wide Regulatory Elements of Gut Tissues in Livestock Species

Zhangyuan Pan¹, Colin Kern², Xiaoqin Xu¹, Ying Wang¹, Michelle M. Halstead¹, Ganrea Chanthavixay¹, Hans Cheng³, Catherine W. Ernst⁴, Pablo J. Ross², Christopher K. Tuggle⁵ and Huaijun Zhou², (1)Animal Science, University of California, Davis, CA, (2)Department of Animal Science, University of California, Davis, CA, (3)USDA, ARS, Avian Disease and Oncology Laboratory, East Lansing, MI, (4)Department of Animal Science, Michigan State University, East Lansing, MI, (5)Animal Science, Iowa State University, Ames, IA

Digestive systems in livestock species play significant roles in animal production and health. Monogastric animals have single stomachs, avian species have gizzards, and ruminants have multi-compartmented stomachs. However, there is limited knowledge on the regulatory elements of digestive systems in livestock species at genome-wide level. Three species, pig, chicken, and cattle, representing these three gut systems, were selected for this study. ChIP-Seq for histone modification marks (H3K4me3, H3K27me3, H3K27ac, and H3K4me1), ATAC-Seq and RNA-Seq were performed to identify genome-wide regulatory elements in six tissues (Stomach/gizzard/rumen, jejunum, duodenum, ileum, colon, and cecum). Current analysis of ChIP-seq data revealed that average peaks of H3K4me3 (32,535), H3K27me3 (92,011), H3K27ac (69,803), and H3K4me1 (91,817) were identified in pig, while 25,567, 85,649, 64,470, 86,943 were identified, respectively, in chickens. Pig using genome-wide read depth seems have more peaks than chicken for all histone marks. However, chicken had significantly higher genome coverage than pig for all marks, which might be due to the chicken's smaller genome. Within chicken, PCA analysis of histone data showed gizzard was clearly separated from intestinal tissues, where duodenum and jejunum usually formed a group separate from a colon/cecum cluster. In pigs, stomach was easily separated from intestinal tissues in all marks, but intestinal tissues had different clustering patterns depending on the mark. Further analysis will identify promoters, enhancers and silencers at genome-wide level. This study will lay a solid foundation for understanding gene regulation of digestion (nutrition), immunity and homeostasis among these distinct gut systems in livestock species.

PO0405: Swine

Identification of Cis-Regulatory Elements and Functional Mutations in the Pig Genome

Yueyuan Xu, Yunxia Zhao, Ye Hou, Xiaolong Qi, Mingyang Hu, Huanhuan Zhou, Yu Luan, Xinyun Li and Shuhong Zhao, Key Laboratory of Agricultural Animal Genetics, Breeding, and Reproduction of the Ministry of Education & Key Laboratory of Swine Genetics and Breeding of Ministry of Agriculture, Huazhong Agricultural University, Wuhan, China

The swine genome sequencing project provides a valuable resource for further improvement of this important livestock specie as meat producer and biomedical model. However, the known positions and function of the *cis*-regulatory elements in the pig genome are so scarce that identification of the functional mutations regulating pig economic traits remains a challenge. To fill this gap, we adopted a strategy similar to ENCODE and Roadmap Epigenomics projects to identify the *cis*-regulatory elements in the pig genome. RNAseq, ATAC-seq, and ChIP-seq for histone markers were performed to generate a comprehensive map of transcriptomes, open chromatin regions, and regulatory elements in a variety of pig tissues from two lean type breeds (Large White and Duroc) and two fatty breeds (Meishan and Enshi Black). We identified over 300,000 *cis*-regulatory elements, representing the most comprehensive functional annotation so far in the pig genome. Among these *cis*-regulatory elements, over 60% of

pig *cis*-regulatory elements are functionally conserved with human *cis*-regulatory elements, which are obviously higher than that between mouse and human. By comparing our Hi-C matrix to human 3D genomic data, we found that over 79% of the boundaries of topologically-associating domains (TADs) were also functionally conserved with human hESC boundaries, indicating the TADs may largely maintain the mammalian ancestral states. It is worth noting that 14 human TADs covers pig inter-chromosome boundaries rearrangements. Genes in these TADs are significantly associated with human head phenotypes. Moreover, based on positions of *cis*-regulatory elements, we efficiently identified 80 mutations in pigs associated with artificial selection. Especially, a G/A mutation in the intron 20 of *IGF1R* was highly associated with the average daily gain of pigs, which was experimentally validated to promote *IGF1R* expression in porcine primary muscle satellite cells. Taken together, our results provide a comprehensive resource of *cis*-regulatory elements in the pig genome, which bridge the gaps of comparative genomics between human and pig genome. Functional mutations identified will be useful targets for pig breeding.

PE0406: Swine

Developmental and Allele-Specific Methylation Patterns in Fetal Liver of Pigs Derived from White Composite x Meishan Reciprocal Crosses

Ryan J. Corbett¹, Haibo Liu², Dan Nonneman³, Timothy P.L. Smith⁴, Nancy E. Raney⁵, Christopher K Tuggle⁶ and Catherine W. Ernst⁵, (1)Genetics & Genome Sciences Graduate Program, Michigan State University, East Lansing, MI, (2)Department of Animal Science, Iowa State University, Ames, IA, (3)USDA, ARS, USMARC, Clay Center, NE, (4)U.S. MEAT ANIMAL RESEARCH CENTER, CLAY CENTER, NE, (5)Department of Animal Science, Michigan State University, East Lansing, MI, (6)Iowa State University, Ames, IA

The liver is a major metabolic organ that influences numerous economically important phenotypes in swine. However developmental and allele-specific gene regulation, including that governed by DNA methylation, are understudied in pig liver. We performed whole-genome bisulfite sequencing (WGBS) of pig fetal liver collected from White Composite and Meishan reciprocal crosses at 30 and 70 days of gestation (dg; N=8,2/stage/cross) in order to assess stage- and allele-specific methylation (ASM). WGBS read alignment and extraction of CpG methylation rates were performed using Bismark, and differential methylation analyses were performed using *methyKit*. We also performed allele-specific mapping of reads and ASM analyses, using SNPs identified from whole-genome sequencing data. Global CpG methylation rates ranged from 59.5-63.6% and were significantly higher in 70dg samples ($p=0.01$). We identified 24,601 differentially methylated regions (DMRs; difference>10%, FDR<1e-5) between stages, 91% of which were hypermethylated at 70dg. DMRs were enriched in gene promoters; 1956 promoter-hypermethylated genes (70dg vs. 30dg) were enriched for GO terms related to early development, while 676 promoter-hypomethylated genes were enriched for lipid and glucose metabolism terms, suggesting decreased and increased transcription of genes involved in these processes, respectively. 529 regions exhibited ASM between White and Meishan alleles, and these were enriched in genes associated with lipid metabolism. Lastly, 430 regions exhibited ASM between maternal and paternal alleles, including regions in the *IGF2* and *IGF2R* gene clusters that are known to exhibit genomic imprinting. This work provides novel insight into epigenetic regulation during pig liver development and has identified genomic regions subject to breed- and parent-specific regulation.

PO0407: Swine

An Integrative GWAS and RNA-Seq Study to Identify SNPs and Transcripts Related to Sperm Quality Traits in Pigs

Marta Godia¹, Antonio Reverter², Rayner González-Prendes³, Yuliaxis Ramayo-Caldas⁴, Anna Castello¹, Joan E. Rodríguez-Gil⁵, Armand Sanchez¹ and Alex Clop^{1,6}, (1)Animal Genomics Group, Centre for Research in Agricultural Genomics (CRAG) CSIC-IRTA-UAB-UB, Bellaterra (Barcelona), Spain, (2)CSIRO Agriculture and Food, St Lucia, QLD, Australia, (3)Wageningen University, Wageningen, Netherlands, (4)IRTA, Caldes de Montbui, Spain, (5)Unit of Animal Reproduction, Department of Animal Medicine and Surgery, Universitat Autònoma de Barcelona, Bellaterra (Barcelona), Spain, (6)Consejo Superior de Investigaciones Científicas (CSIC), Bellaterra, Spain

For the last decades, boars have been selected for their genetic merit on carcass and meat quality traits. However, breeders and researchers are now paying attention to additional phenotypes including sperm quality. The molecular processes affecting sperm quality remain largely unexplored. Genetic pressure in animal breeding is sparking the interest to select for boars with high sperm quality to maximize ejaculate doses and fertility rates. We identified candidate genes, pathways and DNA variants associated to sperm quality in swine by analyzing 25 sperm-related phenotypes with a systems biology approach combining GWAS with 288 boars and genotypes from the Axiom porcine high-density genotyping array and RNA-seq (total and small) from 40 of these pigs. With the GWAS, we identified 12 regions associated to head and neck abnormalities, abnormal acrosomes and motility. Candidate genes included *CHD2*, *KATNAL2* or *SLC14A2*. By RNA-seq, we detected 6,128 significant correlations between sperm traits and gene abundances. To build a robust gene network, only the pair-wise interactions present in both the SNP co-association and the RNA co-abundance networks were kept. The network also included genes which RNA abundances correlated with more than 4 traits. The final network contained genes involved in gamete generation and development, meiotic cell cycle, DNA repair or embryo implantation. A selection of 74 SNPs from the network, GWAS and eGWAS lead hits were used to build a SNP panel that explained between 5 to 36% of the phenotypic variance of these sperm quality traits.

PE0408: Swine

Genetic Diversity of the Swine Leukocyte Antigen Class-II Locus in Infectious Disease Challenges

Kylee Sutton, Lianna Walker, Hiruni R Wijesena, Hiep Vu, Stephen D. Kachman and Daniel Ciobanu, University of Nebraska - Lincoln, Lincoln, NE

Recent genome wide association studies have uncovered QTLs located near the swine leukocyte antigen class-II (*SLAII*) locus (SSC7) which partially explain variation in viremia and immune response following independent challenges with various pathogens. Specifically, a study of two genetic pig lines experimentally infected with PRRSVs revealed a QTL near the *SLAII* region (25 - 27 Mb) accounting for 1.2% of the genetic variation in PRRSVs-specific antibody level in serum. Additionally, following experimental infection with PCV2, a locus near the *SLAII* region (24 - 25 Mb) explained 9.1% of the genetic variation in viral load. Taking in account the known role of *SLAII* in immune response and association with multiple viral pathogens, this locus was characterized in data sets generated by experimental (PRRSV and PCV2) and natural infections (PCV2 and APPV). The diversity of *SLAII* across data sets was based on genotyping by sequencing using reference and novel *DQB1* haplotypes. The specific role of the diversity within the *DQB1* peptide-binding pockets was evaluated across pathogens. Since the commercial genotyping arrays are relatively scarce in SNPs located in the *SLAII* region, a novel Affymetrix SNP array (SowPro90) was designed (103,476 SNPs) that saturated the *SLAII* locus with over 3,100 SNPs and increased SNP density by >100X. This information provides critical knowledge regarding the effective genetic diversity in pigs and the role of *SLAII* in viral disease susceptibility.

PO0409: Swine

Combined Analysis Reveal Association of Changes in Gene Expression and H3K27ac Chromatin Modification at Regulatory Regions in Porcine Alveolar Macrophage in Response to LPS and Poly(I:C)

Juber L. Herrera-Uribe¹, Haibo Liu¹, Kristen Byrne², Zahra Olson², Crystal Loving² and Christopher K Tuggle³, (1)Department of Animal Science, Iowa State University, Ames, IA, (2)USDA-ARS-NADC, Ames, IA, (3)Iowa State University, Ames, IA

Regulation of transcription is associated with changes in chromatin structure by histone modifications (HMs) associated with chromatin accessibility for transcription factors. Alveolar macrophages (AM) play an important role in host defense given that they are the most plastic cells of the immune system, giving them the ability to adapt and to provide an effective immune response against pathogenic microorganisms. By combining RNA sequencing (RNAseq) with chromatin immunoprecipitation and sequencing (ChIP-seq) for four HMs (H3K4me3, H3K4me1, H3K27ac and H3K27me3), we established the chromatin state map of AM, and investigated the potential regulatory effect of these chromatin modifications on RNA changes in AM stimulated with lipopolysaccharide (LPS) and Poly(I:C) at 2h and 6h. The integrative analysis suggests that the differential gene expression between non-stimulated and stimulated AM is significantly associated with changes in H3K27ac at active regulatory regions in

the genome. Although globally changes to chromatin states were minor after stimulations at 2h and/or 6h, we found chromatin state changes for selected differentially expressed genes involved in TLR4, TLR3 and RIG-I signaling pathways. This could suggest that regulatory elements (i.e. active promoters) are already active/poised for immediate inflammatory response in porcine AM. In summary, our data reported here provides the first chromatin state map of AM in response to bacterial and viral mimics, contributing to the Functional Annotation of Animal Genomes (FAANG) project. Furthermore, this work demonstrates the role of HMs, especially H3K27ac, in macrophage response to LPS and Poly(I:C).

PE0410: Swine

Effect of Genotype at a Genetic Marker for GBP5 on Resilience to a Polymicrobial Natural Disease Challenge in Pigs

Ryan L. Jeon¹, Jian Cheng², Austin Putz¹, Qian Dong¹, John Harding³, Michael Dyck⁴, Graham S. Plastow⁵, Frederic Fortin⁶, Joan K. Lunney⁷, Raymond R. Rowland⁸, Bob Kemp⁹ and J.C.M. Dekkers¹, (1)Department of Animal Science, Iowa State University, Ames, IA, (2)Iowa State University, Ames, IA, (3)Department of Large Animal Clinical Sciences, University of Saskatchewan, Saskatoon, SK, Canada, (4)Livestock Gentec, Department of Agricultural, Food and Nutritional Science, University of Alberta, Edmonton, AB, Canada, (5)University of Alberta, Edmonton, AB, Canada, (6)Centre de développement du porc du Québec inc., Québec, QC, Canada, (7)APDL, BARC, ARS, USDA, Beltsville, MD, (8)Department of Diagnostic Medicine/Pathobiology, College of Veterinary Medicine, Kansas State University, Manhattan, KS, (9)PigGen Canada, Guelph, ON, Canada

A genetic marker near the GBP5 gene (WUR0000125, WUR) was previously associated with host response to porcine reproductive and respiratory syndrome (PRRS) virus infection. This study's objectives were to 1) determine whether genotype at WUR is also associated with resilience following a natural polymicrobial disease challenge, 2) investigate the relationship of WUR genotype with its putative causative mutation in GBP5, and 3) compare the association of WUR and the GBP5 mutation with host response to PRRSV infection. Data from two research projects were used: 1) Eight trials of the PRRS Host Genetic Consortium (PHGC) in which ~200 crossbred pigs were infected with the PRRSV2 NVSL 97-7895 to study the effects of genotype at GBP5 and WUR on viral load and weight gain post infection; 2) a natural polymicrobial disease challenge, where 3139 naive crossbred nursery barrows were raised in a nursery-finish barn seeded with pathogens, including PRRSV, to enable expression of disease resilience. Results from the natural disease challenge indicated that the favorable allele for WUR was significantly associated with greater average daily gain ($p=0.04$) and lower numbers of treatments in both the challenge nursery ($p=0.05$) and across the challenge nursery and finisher ($p=0.001$). Results from the PHGC trials indicated that the WUR and GBP5 genotypes were in high but not complete linkage disequilibrium ($r^2=0.94$). However, we were unable to detect a significant difference in the PHGC trials between the association of WUR versus GBP5 with host response to PRRSV ($p=0.23$ and $p=0.46$ for weight gain and viral load). Funded by USDA-NIFA, Genome Canada, Genome Alberta, PigGen Canada.

PO0411: Swine

Searching for Links Between Gut Microbiota Collected Before Vaccination and Variabilities of Vaccine Response in Pigs

Marion Borey¹, Fany Blanc¹, Gaëtan Lemonnier¹, Jean-Jacques Leplat¹, Nicolas Bruneau¹, Deborah Jardet¹, Marie-Noëlle Rossignol¹, Edwige Bouguyon², Laure Ravon³, Fanny Calenge¹, Maria Bernard⁴, Jordi Estellé i Fabrellas¹ and Claire Rogel-Gaillard¹, (1)GABI, INRA, AgroParisTech, Université Paris-Saclay, Jouy-en-Josas, France, (2)VIM, INRA, Université Paris-Saclay, Jouy-en-Josas, France, (3)GenESI, INRA, Surgères, France, (4)INRA, UMR GABI, PSGen-Sigenae, Jouy-en-Josas, France

Understanding and predicting why some animals respond better to vaccination than others is a main concern to strengthen vaccination efficiency. Our aim was to study whether the gut microbiota before vaccination presents composition patterns associated with individual variabilities of vaccine responses in pigs. Ninety-eight Large White piglets were vaccinated against the influenza A virus (IAV) at weaning at 28 days of age (D28) with a booster three weeks later. Stools were collected before the vaccination at D28, and were further processed to perform 16S RNA

gene sequencing (Illumina MiSeq) and assess microbial taxonomic composition. The piglets' humoral response was evaluated by ELISA of seric IAV-specific IgGs and by hemagglutination inhibition assays (HAI) at D49, D56, D63, and D146 to identify extreme animals with either high or low responses to vaccination. Piglets with a richer microbiota had higher levels of HAI at D63 ($p<0.05$) and had a tendency towards more IAV-specific IgGs. Extreme high and low responders for IAV-specific IgGs at D63 had also a dissimilar microbiota ($p<0.01$) and displayed differentially abundant operational taxonomic units (OTUs); bacteria from the Paludibacteraceae family and *Prevotella* genera were more abundant in high responders, while bacteria from *Helicobacter* and *Escherichia-Shigella* genera were more abundant in low responders ($FDR<0.05$). Thus, our results show that the faecal microbiota before vaccination could be further investigated to identify biomarkers predictive of vaccine response levels and analyse the underlying biology.

PE0412: Swine

The Pre-Weaning Gut Microbiota Composition in Piglets and It's Links with Post-Weaning Robustness

Manuel Revilla^{1,2}, Gaëtan Lemonnier¹, Jean-Jacques Leplat¹, Francesca Romana Massacci^{1,3}, Deborah Jarde¹, Marie-Noëlle Rossignol¹, Fany Blanc¹, Marie-José Mercat⁴, Laure Ravon⁵, Yvon Billon⁵, Rafael Muñoz-Tamayo², Nicolas C Friggens², Nathalie Le Floch⁶, Olivier Zemb⁷, Patricia Lepage⁸, Claire Rogel-Gaillard¹ and Jordi Estellé i Fabrellas¹, (1)GABI, INRA, AgroParisTech, Université Paris-Saclay, Jouy-en-Josas, France, (2)MoSAR, INRA, AgroParisTech, Université Paris-Saclay, Paris, France, (3)DISTAL, University of Bologna, Bologna, Italy, (4)IFIP-Institut du porc and Alliance R&D, Le Rheu, France, (5)GenESI, INRA, Surgères, France, (6)PEGASE, INRA, AgroCampus Ouest, Saint-Gilles, France, (7)GenPhySE, INRA, ENVT, Université de Toulouse, Toulouse, France, (8)MICALIS, INRA, AgroParisTech, Université Paris-Saclay, Jouy en Josas, France

The weaning of piglets is often accompanied by health issues like diarrhea. Antibiotics are used for the treatment of diarrhea episodes and also in preventive prophylactic strategies, but there is a strong need for alternatives. The objective of this study is to evaluate the potential of piglet gut microbiota composition analyses to predict individual robustness to weaning. To this end, over 400 Large White piglets were monitored along the weaning period for health and growth traits. In addition to diarrhea episodes, weaning robustness was evaluated with a synthetic phenotype based on the dynamics of body weight, and the population was classified in robust/sensitive according to these criteria. Finally, the gut microbiota composition at weaning (28 days-of-age; w_t0) and one week later (w_t1) was determined by sequencing the bacterial 16S gene in an Illumina MiSeq device. With a sPLS-DA approach using the gut microbiota composition, it was possible to distinguish robust from sensitive animals. Interestingly, a set of 190 OTUs was enough to clearly separate robust from sensitive piglets, with ROC values of 0.92 at w_t0 and of 0.96 at w_t1 time points. Moreover, for this subset of OTUs the correlation between observed and predicted robustness values was 0.84 and 0.88 for w_t0 and w_t1, respectively. Overall, we illustrate here how gut microbiota composition could be used as a relevant source of biomarkers able to predict individual piglet robustness at weaning. In addition, our results could potentially contribute to the identification of antibiotic alternatives at weaning, such as next-generation probiotics

PO0413: Swine

Genome-Wide Association Analysis of Admixed Feral Swine (*Sus scrofa*) Reveals Quantitative Trait Loci Associated with Aujeszky's Disease (Pseudorabies Virus)

Courtney F. Pierce¹, Ryan S. Miller², Scott Speidel¹ and Timothy J. Smyser³, (1)Colorado State University, Fort Collins, CO, (2)USDA-APHIS Center for Epidemiology and Animal Health, Fort Collins, CO, (3)USDA-APHIS National Wildlife Research Center, Fort Collins, CO

Invasive feral swine (*Sus scrofa*) are known to carry pathogens of consequence to humans and domestic animals including Suid herpesvirus 1 (SuHV-1) – the causative agent of Aujeszky's disease (pseudorabies virus [PRV]) – a disease that disrupts the central nervous system, causes respiratory illness, and results in reproductive losses in domestic swine. Contact with infected feral swine presents spillover risks to wildlife, companion animals, domestic livestock and humans. Genetic mechanisms underlying host susceptibility/resistance to Aujeszky's disease are relatively unknown. We sought to identify quantitative trait loci associated with PRV infection among naturally

infected feral swine using a case/control genome-wide association study design. Serology and genotype (55,963 SNP) data were collected on 5,826 feral swine distributed across the invaded range within the United States, of which 1,024 were PRV positive (cases) and 4,802 were negative (controls). Bayesian multiple-SNP regression was conducted for PRV status using GenSel, employing MCMC sampling and Bayes B methodology ($\pi=0.995$). Genetic variance explained by consecutive, non-overlapping 1 MB genomic windows was estimated and those explaining >1% of the genetic variance were considered significant. Genomic windows associated with PRV were identified on chromosome 10, 8, and 13, accounting for 1.86%, 1.46% and 1.03% of the genetic variance, respectively. Our results align with previous work that identified candidate genes associated with clinical disease following experimental challenge with PRV in domestic pigs. Using genetic tools to quantify variation in disease susceptibility among feral swine populations enables improved identification of drivers of disease and the concomitant spillover risk to domestic herds.

PE0414: Swine

Anthrax Toxin Receptor 1-Knockout Pigs are Protected from Senecavirus A Infection

Paula R. Chen¹, Raymond R. Rowland², Vlad Petrovan², Maureen Sheahan², Chanran Ganta³, Kristin M. Whitworth¹, Melissa S. Samuel¹, Joshua A. Benne¹, Raissa F. Cecil¹, Lee D. Spate¹, Kelly Lechtenberg⁴, Kevin D. Wells¹ and Randall S. Prather¹, (1)Division of Animal Sciences, University of Missouri, Columbia, MO, (2)Department of Diagnostic Medicine/Pathobiology, College of Veterinary Medicine, Kansas State University, Manhattan, KS, (3)Kansas State Veterinary Diagnostic Laboratory, College of Veterinary Medicine, Kansas State University, Manhattan, KS, (4)Midwest Veterinary Services, Inc., Oakland, NE

Senecavirus A (SVA) has been the cause of numerous cases of vesicular disease in swine across the world in recent years. Studies investigating the oncolytic properties of SVA in humans revealed anthrax toxin receptor 1 (ANTXR1) as its probable receptor. The objective of the current study was to determine if ANTXR1 functioned as the receptor for SVA in pigs by employing the CRISPR/Cas9 system to edit exon 1 and create a premature stop codon. Two founder *ANTXR1*-knockout pigs and two age-matched wild-type pigs were challenged with SVA. Serum, fecal swabs, and nasal swabs were collected throughout the duration of the study. Presence of viral nucleic acid was determined by PCR, and SVA antibody responses were assessed. *ANTXR1*-knockout pigs had a distinct phenotype, including frontal bossing and wide, short statures, which is characteristic of GAPO syndrome in humans. The knockout pigs did not develop vesicular lesions while the wild-type pigs had coronary band lesions after SVA infection. Moreover, SVA nucleic acid was not detected in either *ANTXR1*-knockout pig, but virus was present in fecal and nasal swabs of one knockout pig. The same pig demonstrated evidence for production of SVA-specific antibodies; however, both knockout pigs did not exhibit virus neutralizing activity. Because founder pigs created by microinjection of the CRISPR/Cas9 system can have mosaic genotypes, a study on F1s is warranted. Overall, knocking out *ANTXR1* appears to confer protection against SVA infection in pigs, and modulation of this region may be needed to correct the phenotype associated with the edit.

PO0415: Swine

Gene Editing and Signatures of Adaptive Evolution Emphasize the Role of Synaptogyrin-2 in PCV2b Susceptibility

Lianna Walker¹, Hiep Vu¹, Kristi Montooth¹, Tudor Borza² and Daniel Ciobanu¹, (1)University of Nebraska - Lincoln, Lincoln, NE, (2)Dalhousie University, Truro, NS, Canada

Porcine circovirus 2 (PCV2) has been identified as the causative pathogen of a set of syndromes collectively referred to as Porcine Circovirus Associated Diseases (PCVAD). However, infection with this small single stranded DNA virus does not guarantee disease. Observed variation in incidence and severity of PCVAD between breeds suggests an important role of host genetics in PCV2 susceptibility. A large-scale genome-wide association study of ~1,000 pigs revealed a prominent quantitative trait locus (QTL) for PCV2b viral load located on the proximal end of chromosome 12. Further investigation of this QTL region uncovered a missense mutation within the synaptogyrin-2 gene (*SYNGR2* p.Arg63Cys) associated with PCV2b viral load. This polymorphism is located within the second exon of *SYNGR2*, which encodes the first intraluminal loop and a critical domain necessary for protein function. CRISPR-Cas9 mediated gene editing of PK15 cells focused solely on the second exon of *SYNGR2* has provided

direct evidence for the involvement of this gene in PCV2b infection. Sequence analyses using the ratio of non-synonymous to synonymous mutations (dN/dS) indicated overall negative selection acting to conserve the *SYNGR2* protein sequence. Interestingly, multiple positions within the intraluminal loops were identified as possible sites of adaptive evolution across mammalian species. These findings directly link *SYNGR2* to PCV2b susceptibility and highlight the intraluminal loops as sites of potential adaptive evolution to environmental cues, including host-pathogen interactions.

PE0416: Swine

Differentially Expressed Immune Genes Identified in the Placenta and Fetal Thymus in Fetuses of Pregnant Gilts Infected With PRRS Virus

Angelica Van Goor¹, Kristen Walker¹, Alex Pasternak², Carolina Malgarin³, Daniel J. MacPhee⁴, John Harding⁵ and Joan K. Lunney⁶, (1)USDA ARS NEA BARC, Beltsville, MD, (2)Department of Animal Science, Purdue University, West Lafayette, IN, (3)Dept. Large An Clin Sci., University of Saskatchewan, Saskatoon, SK, Canada, (4)Department of Veterinary Biomedical Sciences, University of Saskatchewan, Saskatoon, SK, Canada, (5)Department of Large Animal Clinical Sciences, University of Saskatchewan, Saskatoon, SK, Canada, (6)APDL, BARC, ARS, USDA, Beltsville, MD

Our goal is to identify factors controlling maternal-fetal transmission of porcine reproductive and respiratory syndrome (PRRS) virus. Pregnant gilts at 85 days gestation were infected ($n = 31$) with PRRS virus (NVSL97) or mock infected ($n = 7$). At 12 days post infection gilts were euthanized and fetal preservation status determined. Maternal fetal interface (MFI) samples were carefully dissected so that placental (PLC) viral load (VL) could be independently assessed along with fetal thymus (THY) and serum. Fetuses were grouped by preservation status and PRRS VL: control (CTRL), no virus detected (UNIF), virus detected in the placenta only with viable (PLCO-VIA) or meconium-stained fetus (PLCO-MEC), low viral load with viable (LVL-VIA) or meconium-stained fetus (LVL-MEC), and high viral load with viable (HVL-VIA) or meconium-stained fetus (HVL-MEC). Tissue RNA was extracted and differentially expressed genes (DEG) evaluated using a 286 gene NanoString array (designed on biomarkers previously predicted to alter PRRS resistance/ susceptibility). Statistical analyses were performed using JMP SAS software for ANOVA with Tukey's multiple testing correction ($P \leq 0.05$). No DEG were identified in either tissue for UNINF or PLCO-VIA compared to CTRL. We identified 19, 6, 63, 62, and 88 DEG for PLC, and 1, 6, 91, 52, and 47 DEG for THY, in PLC-MEC, LVL-VIA, LVL-MEC, HVL-VIA, and HVL-MEC compared to CTRL, respectively. Preliminary analysis of these data suggests transcriptional dysregulation of PRRSV response genes is correlated with fetal rather than placental infection. The unique gene expression patterns identified may help breed for PRRS resistant pigs.

PO0417: Swine

Transcriptome Analysis of Host Resistance to PRRS in Tongcheng \times Large White Advanced Generation Inter-Cross Population.

Xiang Zhou¹, Yuan Wang¹, Xiangge Meng¹, Bang Liu¹ and Huazhong Agricultural University, (1)Huazhong Agricultural University, Wuhan, China

Porcine reproductive and respiratory syndrome (PRRS), caused by PRRS virus (PRRSV), is one of the most important infectious diseases in pig industry. Disease-resistance breeding has become an effective approach to prevent and control PRRS epidemic. Our previous studies have showed that Tongcheng (TC) pigs, a Chinese local breed, display stronger resistance or tolerance to PRRSV. In order to explore genetic resistance to PRRSV, we constructed a Tongcheng \times Large White advanced generation inter-cross population. A total of 122 piglets were artificial infected with highly pathogenic PRRSV WUH3 in vivo. Blood samples and Body weight were collected at 0, 4, 7, 11, 14, 21, 28 and 35 days post infection (DPI). Viral load and weight gain were measured individually up to 35 DPI. Furthermore, PRRS-resistant pigs and PRRS-susceptible pigs were selected by viral load and weight gain. Then, transcriptome profiling of white cells derived from PRRS-resistant pigs and PRRS-susceptible pigs were performed using RNA-seq at 0, 4, 7, 11 DPI. For PRRS-resistant pigs, 1040 genes were differently expressed in response to PRRSV infection. While, 881 differently expressed genes (DEGs) were identified in PRRS-susceptible pigs. There are only 235 genes are common DEGs between PRRS-resistant pigs and PRRS-susceptible pigs. Pathway enrichment of differentially expressed genes (DEGs) confirmed that immune response, inflammatory

response, cytokine-cytokine receptor interaction pathways play important roles for PRRSV clearance. This study provides us new insight into the transcriptomic comparison of PRRS-resistant pigs and PRRS-susceptible Tongcheng × Large advanced generation inter-cross population.

PE0418: Swine

Protein Levels in Blood of Young Healthy Pigs As Indicators of Disease Resilience

Yulu Chen¹, Steven M Lonergan², Kyu-Sang Lim¹, Austin Putz³, Micheal Dyck⁴, Frederic Fortin⁵, John Harding⁶, Graham S. Plastow⁴ and Jack C.M. Dekkers¹, (1)Iowa State University, Ames, IA, (2)Department of Animal Science, Iowa State University, Ames, IA, (3)Hypor, Ames, IA, (4)University of Alberta, Edmonton, AB, Canada, (5)Centre de développement du porc du Québec inc., Québec, QC, Canada, (6)Department of Large Animal Clinical Sciences, University of Saskatchewan, Saskatoon, SK, Canada

Disease resilience is the ability to maintain performance under pathogen exposure. Selection for disease resilience is difficult because nucleus breeding populations must be kept in a high health environment. Biomarkers for disease resilience that can be measured at an early age in high-health conditions could overcome this limitation. Proteins act as a working force to determine the organism's phenotype and the blood proteome has been used to identify biomarkers for some human diseases. Our objective was to explore the blood proteome of young healthy pigs for potential biomarkers of disease resilience. Seven batches of 60-75 healthy weaned Yorkshire x Landrace barrows (n=405) were entered into a quarantine nursery and blood samples were collected around 27 days of age. One week later, the pigs were moved to a nearby natural challenge facility, which was established by seeding the barn with pigs that were naturally infected with multiple pathogens and maintained using a continuous flow system. The levels of 481 proteins were quantified in the blood samples using the Tandem Mass Tag based LC-MS/MS method. Associations of the protein levels with performance and disease resilience were evaluated using mixed linear models and several proteins showed suggestively significant associations ($p < 0.05$). The abundance of 95 proteins was found to be moderately heritable ($h^2 > 0.1$). Estimation of genetic correlations of protein abundance with resilience is underway. In conclusion, protein levels in blood of young healthy pigs show potential as biomarkers for disease resilience. Funding from USDA-NIFA, Genome Canada, PigGen Canada.

PO0419: Swine

Quantitative Genetic Analysis of the Blood Transcriptome of Young Healthy Pigs to Improve Disease Resilience

Kyu-Sang Lim¹, Jian Cheng¹, Christopher K Tuggle¹, Michael Dyck², Frederic Fortin³, John Harding⁴, Graham S. Plastow⁵, Jack C.M. Dekkers¹ and PigGen Canada, (1)Iowa State University, Ames, IA, (2)Livestock Gentec, Department of Agricultural, Food and Nutritional Science, University of Alberta, Edmonton, AB, Canada, (3)Centre de développement du porc du Québec inc., Québec, QC, Canada, (4)Department of Large Animal Clinical Sciences, University of Saskatchewan, Saskatoon, SK, Canada, (5)University of Alberta, Edmonton, AB, Canada

The complexity of gene expression is determined by not only the environment but also genetics. Here, we estimated the heritability of gene expression in the blood of young healthy piglets (~27 days of age) and its genetic correlation with measures of resilience after exposure to a natural polymicrobial disease challenge. Weaned barrows (n=3,205, Yorkshire x Landrace, in 50 batches) from healthy multiplier farms were evaluated for disease resilience in an experimental facility consisting of a high-health quarantine nursery and a challenge nursery and finisher. All pigs were genotyped with the 650k porcine genotyping array. Gene expression in blood samples collected in the quarantine nursery (n=903) was quantified by 3'mRNA sequencing. Average daily gain (ADG) over each stage for pigs that survived was evaluated as a resilience phenotype (qNurADG, n=3,138; cNurADG, n=2,784; cFinADG, n=2,341). Heritability estimates for qNurADG, cNurADG and cFinADG were 0.32 (± 0.04), 0.25 (± 0.04), 0.29 (± 0.05), respectively. Among 15,872 evaluated genes, the expression of 292 and 1,487 genes had a high ($h^2 > 0.4$) and moderate ($0.4 > h^2 \geq 0.2$) estimates of heritability, respectively. The top 5,006 heritable genes were used to estimate genetic correlations with ADG. The numbers of genes that showed significant ($p < 0.05$) genetic correlations with qNurADG, cNurADG, and cFinADG were 288, 199, and 191, respectively. These results provide new insight into the heritability of the porcine blood transcriptome, implicating its possible use in young healthy pigs as early

predictors to improve disease resilience. Funding from USDA-NIFA, Genome Canada, Genome Alberta, and PigGen Canada.

PE0420: Swine

Combined Analysis Reveal Association of Changes in Gene Expression and H3K27ac Chromatin Modification at Regulatory Regions in Porcine Alveolar Macrophage in Response to LPS and Poly(I:C)

Juber L. Herrera-Urbe, Department of Animal Science, Iowa State University, Ames, IA

Regulation of transcription is associated with changes in chromatin structure by histone modifications (HMs) associated with chromatin accessibility for transcription factors. Alveolar macrophages (AM) play an important role in host defense given that they are the most plastic cells of the immune system, giving them the ability to adapt and to provide an effective immune response against pathogenic microorganisms. By combining RNA sequencing (RNAseq) with chromatin immunoprecipitation and sequencing (ChIP-seq) for four HMs (H3K4me3, H3K4me1, H3K27ac and H3K27me3), we established the chromatin state map of AM, and investigated the potential regulatory effect of these chromatin modifications on RNA changes in AM stimulated with lipopolysaccharide (LPS) and Poly(I:C) at 2h and 6h. The integrative analysis suggests that the differential gene expression between non-stimulated and stimulated AM is significantly associated with changes in H3K27ac at active regulatory regions in the genome. Although globally changes to chromatin states were minor after stimulations at 2h and/or 6h, we found chromatin state changes for selected differentially expressed genes involved in TLR4, TLR3 and RIG-I signaling pathways. This could suggest that regulatory elements (i.e. active promoters) are already active/poised for immediate inflammatory response in porcine AM. In summary, our data reported here provides the first chromatin state map of AM in response to bacterial and viral mimics, contributing to the Functional Annotation of Animal Genomes (FAANG) project. Furthermore, this work demonstrates the role of HMs, especially H3K27ac, in macrophage response to LPS and Poly(I:C).

PO0421: Swine

Exploring Complete Blood Count Data As a Disease Resilience Phenotype in Pigs from a Natural Disease Challenge Model

Xuechun Bai¹, Tianfu Yang², Austin M. Putz³, Zhiquan Wang¹, Frederic Fortin⁴, John C. S. Harding⁵, PigGen Canada⁶, Michael K. Dyck¹, Jack C. M. Dekkers³, Catherine J. Field¹ and Graham S. Plastow¹, (1)Livestock Gentec, Department of Agricultural, Food and Nutritional Science, University of Alberta, Edmonton, AB, Canada, (2)Livestock Gentec, University of Alberta, Canada, Edmonton, AB, Canada, (3)Department of Animal Science, Iowa State University, Ames, IA, (4)Centre de développement du porc du Québec inc., Québec, QC, Canada, (5)Department of Large Animal Clinical Sciences, University of Saskatchewan, Saskatoon, SK, Canada, (6)PigGen Canada, Guelph, ON, Canada

Disease resilient animals maintain their performance in the face of infection, and it is anticipated that genetic improvement of resilience will sustainably increase production efficiency. The objective of this study was to identify phenotypes and genomic regions related to disease resilience using complete blood count (CBC) from a natural disease challenge model, established to mimic the disease pressure at the commercial level of pig production. Pigs were classified into resilient, average, susceptible, and dead groups based on their divergent responses to the challenge. Blood samples for CBC were drawn at 2-weeks before, and at 2 and 6-weeks after exposure to the challenge; Blood 1, Blood 3 and Blood 4, respectively. Resilient animals were found to be primed to initiate a faster adaptive immune response and recover earlier from infection, with greater increases in lymphocyte concentration from Blood 1 to Blood 3 and for hemoglobin from Blood 3 to Blood 4, but a lower neutrophil concentration in Blood 4 among groups (FDR<0.05). CBC traits in response to the challenge were moderately heritable ($h^2>0.1$) and genetically correlated with the resilience traits of growth and treatment incidence (-0.82 to 0.89). Various genomic regions that were significantly associated with CBC traits are located within or nearby genes with potential roles in hematopoiesis, granulopoiesis and granulocytic differentiation, erythroid and megakaryocytic differentiation, inflammatory response, antigen presentation, and regulation of immune responses. These results provide support for the use of CBC traits as phenotypes in commercial systems as part of developing predictions for disease resilience.

PE0422: Swine

Protein Levels in Blood of Young Healthy Pigs as Indicators of Disease Resilience

Yulu Chen, Iowa State University, Ames, IA

Disease resilience is the ability to maintain performance under pathogen exposure. Selection for disease resilience is difficult because nucleus breeding populations must be kept in a high health environment. Biomarkers for disease resilience that can be measured at an early age in high-health conditions could overcome this limitation. Proteins act as a working force to determine the organism's phenotype and the blood proteome has been used to identify biomarkers for some human diseases. Our objective was to explore the blood proteome of young healthy pigs for potential biomarkers of disease resilience. Seven batches of 60-75 healthy weaned Yorkshire x Landrace barrows (n=405) were entered into a quarantine nursery and blood samples were collected around 27 days of age. One week later, the pigs were moved to a nearby natural challenge facility, which was established by seeding the barn with pigs that were naturally infected with multiple pathogens and maintained using a continuous flow system. The levels of 481 proteins were quantified in the blood samples using the Tandem Mass Tag based LC-MS/MS method. Associations of the protein levels with performance and disease resilience were evaluated using mixed linear models and several proteins showed suggestively significant associations ($p < 0.05$). The abundance of 95 proteins was found to be moderately heritable ($h^2 > 0.1$). Estimation of genetic correlations of protein abundance with resilience is underway. In conclusion, protein levels in blood of young healthy pigs show potential as biomarkers for disease resilience. Funding from USDA-NIFA, Genome Canada, PigGen Canada.

P00423: Swine

Fine Mapping Genetic Variants Associated with Age at Puberty and Sow Fertility Using SowPro90 Genotyping Array

Hiruni R Wijesena, Matt L. Spangler, Stephen D. Kachman and Daniel Ciobanu, University of Nebraska - Lincoln, Lincoln, NE

Gilts that reach puberty earlier have a greater probability of having more lifetime litters. Identifying pleiotropic genetic variants associated with age at puberty (AP) and other fertility traits expressed late in life, such as reproductive longevity, has the potential to improve the accuracy of genomic prediction for sow reproductive traits. A novel Affymetrix SNP array, 'SowPro90' was developed saturating the major QTLs for AP with SNPs located in genes overlapping these regions. Maternal UNL crossbred sows representing the tails of the distribution for Porcine SNP60 BeadArray (53,529 SNPs) derived genomic prediction values for AP (early, n=147; late, n=123) were genotyped with SowPro90 to fine map pleiotropic polymorphisms associated with AP and reproductive longevity. A reference population including a subset of UNL crossbred sows and sows from the parental genetic lines genotyped with SowPro90 (86,867 SNPs) were used to infer SowPro90 haplotypes to the entire UNL population (n=1,441) previously genotyped with SNP60 BeadArray. A haplotype-based genome-wide association (n=1,931) identified major QTL regions for AP (SSC2, 13-14Mb, SSC7, 83-85Mb, SSC14, 43-45Mb, and SSC18, 37-38Mb). The top pleiotropic SNP on SSC7 (83.3Mb, $P=0.0003$) explained 2.9% of the phenotypic variation of AP and 1.1% of the phenotypic variation of lifetime number of parities produced. As the number of favorable alleles for this SNP increased, AP decreased by 5 days and lifetime number of parities increased by 0.26. A candidate gene (*NR2F2*) located upstream of this SNP (~280Kb) is implicated in progesterone receptor signaling during embryonic implantation and litter size in pigs.

P00425: Aquaculture

A Long Reads-Based Trio-Binning De-Novo Assembly of the North American Atlantic Salmon Genome

Guangtu Gao¹, Geoff Waldbieser², Ramey C. Youngblood³, Michael Pietrak⁴, Brian E. Scheffler⁵, Caird E. Rexroad III⁶, Brian Peterson⁴ and Yniv Palti¹, (1)USDA-ARS-NCCCWA, Kearneysville, WV, (2)USDA-ARS Warmwater Aquaculture Research Unit, Stoneville, MS, (3)Mississippi State University, Mississippi State, MS, (4)National Cold Water Marine Aquaculture Center, ARS-USDA, Orono, ME, (5)USDA-ARS, Stoneville, MS, (6)USDA, Beltsville, MD

A high-quality reference genome assembly is available for the European sub-species of Atlantic salmon. Given the genomic differences between European and North American (NA) Atlantic salmon, another reference de-novo assembly is needed for the NA Atlantic salmon. Currently, we do not have a homozygous salmon of NA origin for a de-novo assembly, but this problem may be overcome by the recently published trio-binning assembly approach. Trio-binning uses short Illumina reads from the two parental genomes to partition the long reads obtained from the heterozygous offspring into haplotype-specific sets. Each haplotype is then assembled independently to reconstruct the two parental genomes. To accomplish this, we initially generated 104x genome coverage in PacBio Sequel long-read sequence from a single salmon male (Chromosome 3/6 Y lineage) from the St. John River broodstock of the USDA breeding program in Maine. We then generated over 40x Illumina paired-end reads from each parent and used the short reads to separate the maternal and paternal long reads. For each parental genome, contigs were assembled from the pre-selected long reads using the Canu pipeline and consensus sequence was error-corrected using two iterations of Arrow with the PacBio raw reads. The Canu assembly contained 12,416 and 13,021 contigs with an N50 contig length of 768,218bp and 796,316bp for the female and male parental haplotypes, respectively. The total lengths of the female and male assemblies were 3.32Gb and 3.31Gb, respectively. A BUSCO analysis detected 94.1% and 93.1% of conserved Actinopterygii genes in the female and male assembly, respectively. We are currently adding over 50x genome coverage with PacBio long reads from the heterozygous offspring to further improve the contiguity of the two parental assemblies. The two assemblies will also be further improved with a Bionano optical map and Hi-C proximity ligation sequence data to produce super-scaffolds and correct mis-joined scaffolds.

PE0426: Aquaculture

Inbreeding and Runs of Homozygosity in Farmed Coho Salmon Populations

Grazyella M. Yoshida and José M Yáñez, University of Chile, Santiago, Chile

We performed runs of homozygosity (ROH) analysis to identify and characterize ROH patterns in three farmed coho salmon populations and compared the estimates of inbreeding coefficient calculated from the run of homozygosity (FROH) genomic relationship matrix (FGRM) and pedigree-based relationship matrix (FPED). We used a total of 240 individuals from independent population A (POP A) and B (POP B), and a population resulted from crossed both POP A and B (POP C). All animals were genotyped using 200K Affymetrix Axiom myDesign Custom Array and a total of 102,129 markers passed in quality control. We identified ROH in all animals in the three coho salmon populations, totaling 3,250, 1,605 and 273 for POP A, B and C, respectively. The higher number of ROHs were identified in the first and second (1-2Mb and 2-4 Mb) ROH length classes, expected to correspond to the reference ancestral population dating to 50 and 20 generations ago, respectively. For most of animals from POP A and B was found long ROHs segments (>16 Mb), suggested that the individuals are about six generations separated from common ancestor that contributed with the IBD fragment, whereas for POP C segments longer than 8 Mb were found just for few animals, indicating that the recent admixture was effective in broke down the ROH segments. Moreover, the inbreeding coefficient estimated using genomics or pedigree methods can have varied among populations and the high correlations between genomics inbreeding suggested that these are the more accurate methods to estimating the inbreeding.

PO0427: Aquaculture

DNA Methylation Dynamics in Atlantic Salmon (*Salmo salar*) after Being Challenged with High Temperature and Moderate Hypoxia

Anne Beemelmans, Department of Ocean Sciences, Memorial University of Newfoundland, St. John's, NF, Canada

The marine environment is predicted to become warmer, and more hypoxic, over this century and these conditions may become a challenge for cultured Atlantic salmon by negatively affecting their growth, immunology and welfare. DNA methylation mediates phenotypically plastic responses in gene expression and can potentially facilitate acclimatization responses. Thus, we measured DNA methylation from salmon that were subjected to: *i*) control conditions (normoxia, 12°C) throughout the experiment; *ii*) an incremental increase in temperature (12°C to 20°C, at 1°C per week) and then held at 20°C for 4 weeks; or *iii*) the former temperature regimen in combination with moderate hypoxia (~70% air saturation). DNA methylation levels were measured at CpG sites within a ~500 bp region (Promotor, 5'UTR, Exon, Intron) of six important liver biomarker genes (*cribp*, *jund*, *pkd3*, *prdx6*, *serpinh1*,

and *ucp2*). Considering both experimental groups, we found 12 CpGs (out of 94 total) across the six genes that were differentially methylated when exposed to 20°C for 3 days, whereas only 6 CpGs from three genes (*jund*, *prdx6* and *ucp2*) were affected after 4 weeks at 20°C. At both time points, the three treatment groups could be clearly distinguished over a differential clustering structure based on the methylation levels of all significantly affected CpGs. Further, we report significant relationships between CpG methylation and the mRNA expression of these genes that are complex and dynamic. These changes in DNA methylation might be an important regulatory mechanism allowing Atlantic salmon to quickly respond to new environmental changes associated with global warming.

PE0428: Aquaculture

The King 8: A SNP Based Test for Continent of Origin in Atlantic Salmon

Melissa Allen¹, John Buchanan², Jason Stannard² and America Fujimoto², (1)Center For Aquaculture Technologies, Souris, PE, Canada, (2)Center for Aquaculture Technologies, San Diego, CA

The potential for aquaculture escapees to compromise the genetic integrity of local wild stocks is a hot topic issue that has significant implications for the Atlantic salmon aquaculture industry. Evidence of European Atlantic salmon genetic markers being found amongst wild North American stocks has prompted regulatory agencies in both Canada and the US to enforce mandatory screening of commercial broodstock to genetically confirm North American origin. Currently, testing for continent of origin (COO) relies on genotyping 7 microsatellites (“The King 7”) developed to discern between North American and European Atlantic salmon stocks. However, microsatellite markers are known to be more error prone and less reproducible relative to genetic markers such as single nucleotide polymorphisms (SNPs). Given that SNP technology translates easily between laboratories, the high degree of reliability and repeatability of SNPs compared to microsatellites make them excellent candidates for long-term screening of COO. Here, we present a panel of 8 SNPs (“The King 8”) designed to assign COO with at least the same level of accuracy as the traditional King 7 microsatellite panel. The SNP panel was validated among North American reference populations (n = 425) from Maine, Quebec, PEI and Nova Scotia as well as European reference populations (n = 315) from Norway, Scotland and Sweden. Transitioning COO screening to SNP based tests is a logical next step for the industry as these SNPs could be easily integrated into pre-existing SNP panels for applications such as pedigree assignment, creating more efficient genomic tools for producers.

PO0429: Aquaculture

DNA Methylation and Transcriptomic Changes Involved in Atlantic Salmon Sexual Maturation

James Kijas, CSIRO Agriculture, Brisbane, Australia, Antonio Reverter, CSIRO Agriculture Flagship, Brisbane, Australia, Bradley Evans, Tassal, Hobart, TAS, Australia, Harry King, CSIRO, Hobart, TAS, Australia and Amin Mohamed, CSIRO, Brisbane, QLD, Australia

Atlantic salmon farming promotes growth in conditions which mean animals may complete sexual development at weights below harvest size. This can lead to a reduction in productivity and prompted us to investigate the biological mechanisms that control the timing of sexual maturation. We performed a time course experiment, whereby animals were manipulated with photoperiod before tissues were collected across the time window when animals commence sexual development. We performed whole genome bisulfite sequencing of three salmon tissues (pituitary, ovary and liver) at both the beginning and end of the experiment, to take a first look at the patterns of DNA methylation and examine how they change in response to the onset of an important life history trait. Comparison across timepoints revealed 6,373 differentially methylated regions (DMRs), of which approximately 50% were located within genes (DMGs). The ovary underwent the most profound remodelling, with a strong bias towards increased methylation levels (hyper-methylation). We also performed deep transcriptomic profiling (RNA-seq) of the same tissues to explore the relationship between methylation changes and gene expression. Weak correlation was observed considering all available genes, suggesting methylation may not be the key epigenomic regulator of global expression in the context of our experiment. However, we found a significant overlap between DMGs and differentially expressed genes in the ovary. Taken together, our results suggest chromatin remodelling genes play a role in the commitment of animals to the sexual maturation pathway. They also open the way for the identification of functional variants that can be used in advanced breeding approaches to boost productivity in Atlantic salmon farming.

PE0430: Aquaculture

High-Resolution Phylogenetic Gut Microbial Community Profiling of Farmed Chinook Salmon (*Oncorhynchus tshawytscha*)

Ruixiang Zhao, University of Tasmania, Hobart, TAS, Australia

To investigate the relationship between Chinook salmon gut microbiota and fish health, detailed classification and characterization of the gut microbial community in relation to health, diet and environmental factors is essential. We present here the first in-depth phylogenetic profiling of the gut microbial community of farmed Chinook salmon (*Oncorhynchus tshawytscha*), based on high-throughput sequencing of the 16S rRNA genes V1-V3 amplicons. Faecal samples were collected from Chinook salmon raised to market size in freshwater and marine farms in New Zealand over an 18-month period across a standard commercial production cycle. Gastrointestinal microbial community structures were highly dynamic but similar among individual fish specimens within each farm. In freshwater and saltwater salmon, the gut microbial communities were dominated by Proteobacteria, whereas Firmicutes showed the highest relative prevalence in freshwater salmon. Species richness and diversity were significantly higher in freshwater salmon than saltwater salmon. Other abiotic factors (water temperature and geographical location) were also investigated for both salinity groups, but the individual effect of these factors was minimal. Biotic factors (fish life stage and fish size) had more significant effects on the beta diversity of gut microbiota in both freshwater and saltwater habitats. Our study provides a detailed description of the gut microbial community of farmed Chinook salmon and shows that besides the impact of salinity, fish biotic effects are more influential than farm biogeography. This will likely contribute to a greater understanding of the true functional interaction between the fish and the microbial community of the gut.

PO0431: Aquaculture

Impacts of Functional Feed Ingredients on Mucosal Immunity and Microbiota of Atlantic Salmon and Potential Implications for Sea Lice Resistance

Jacob W Bledsoe¹, Mike Pietrak², Gary Burr², Brian Peterson² and Brian C Small¹, (1)University of Idaho, Hagerman, ID, (2)National Cold Water Marine Aquaculture Center, ARS-USDA, Franklin, ME

Aquaculture production of Atlantic salmon is worth approximately US\$10.5 billion globally and one of the greatest challenges impacting salmon production is losses brought on by ectoparasitic infections from the salmon louse *Lepeophtheirus salmonis* (i.e. sea lice). Historically, sea lice infections were managed by chemical therapeutics, though recently sea lice have shown resistance to most approved drugs. There is now much interest in alternative treatments such as functional feeds, which are thought to alter the host's inflammatory response to louse infection or possibly interfere with chemical signaling required for louse virulence. A twelve-week feeding trial was conducted at the National Cold Water Marine Aquaculture Center (ARS-USDA) with post-smolt Atlantic salmon (358 g \pm 17 initial) to evaluate effects of functional feeds on growth performance, microbiota, mucosal immunity, and sea lice resistance. Treatments included: (1) Control Diet (Ctrl), (2) Ctrl + Coconut oil (98% lipid replacement) (CO), (3) Ctrl + 0.4% mannan oligosaccharides (MOS), and (4) Ctrl + Coconut oil + MOS (CO-MOS). Fish growth was measured at six and twelve weeks, and all diets showed acceptable growth with no significant differences by dietary treatment. Following the feeding trial, the gut, gill, and skin microbiota were sampled from three fish tank⁻¹ ($N = 18$ diet⁻¹) and characterized by V3V4-16S rRNA gene sequencing. In addition, parallel tissue samples and peripheral blood leukocytes were also collected, and quantitative gene-expression analysis was performed on key host immune biomarkers. Remaining fish ($N = 72$ diet⁻¹) were then subjected to common-garden sea lice challenges (4-hour static bath; 100 copepodids *L. salmonis* fish⁻¹) followed by 14 days of observation for abnormality or mortality, with no significant impact of diet observed on final lice infection density. Microbiota composition and host gene-expression results will be presented with comparisons across the three sampled mucosal body sites, as well as comparisons by dietary treatment.

PE0432: Aquaculture

A Long Reads-Based De-Novo Assembly of the Rainbow Trout Arlee-Line Genome

Guangtu Gao¹, Geoff Waldbieser², Ramey C. Youngblood³, Paul A. Wheeler⁴, Brian E. Scheffler⁵, Gary H. Thorgaard⁶ and Yniv Palti¹, (1)USDA-ARS-NCCCWA, Kearneysville, WV, (2)USDA-ARS

Warmwater Aquaculture Research Unit, Stoneville, MS, (3)Mississippi State University, Mississippi State, MS, (4)Washington State University, Pullman, WA, (5)USDA-ARS, Stoneville, MS, (6)School of Biological Sciences and Center for Reproductive Biology, Washington State University, Pullman, WA

Although the most recent version of the rainbow trout genome assembly from the Swanson line has greatly improved the genome reference and is reliable for genes' prediction, it contains 420,055 spanned gaps and 7,839 unspanned gaps (GCA_002163495.1). Hence, there is still a need to improve the contiguity and completeness of the reference assembly, which is now possible with long-read DNA sequencing technologies. Currently, we are also working towards generating a rainbow trout "pan-genome" reference that will better represent the genetic diversity in this species. The Arlee doubled haploid YY male line has a different genetic background from the Swanson line. It was originated from a domesticated strain that was originally collected from the northern California coast. For the Arlee genome assembly, we generated 111x genome coverage in long-read sequence data using the PacBio Sequel system. The read length distribution has N50 of ~33 kb and an average read length greater than 20 kb. Contigs were assembled using the Canu pipeline and consensus sequence was error-corrected using two iterations of Arrow with the PacBio reads followed by one iteration of Freebayes using Illumina paired-end reads. The Canu assembly contained 1,591 contigs with an N50 contig length of 9,835,815 bp, which is a major improvement in contiguity compared to the current Swanson assembly. The assembly was further improved with a Bionano optical map and Hi-C proximity ligation sequence data to produce super-scaffolds and correct mis-joined scaffolds. This improved the assembly to a total of 2.34 Gb in 919 scaffolds with an N50 length of 47,542,702 bp. The range of the scaffolds' length distribution after Bionano and Hi-C was 16,956 bp – 90,526,592 bp. A BUSCO analysis detected 96.6% of conserved Actinopterygii gene content in this assembly. We are currently using the rainbow trout high-density genetic map to guide chromosomal alignment of scaffolds.

PO0433: Aquaculture

Heritability Estimates and Genome-Wide Association Analysis of Egg Quality Traits in Commercial Rainbow Trout Breeding Populations

Yniv Palti¹, Roger L. Vallejo², Guangtu Gao³, Roseanna L. Long⁴, Kristy L. Shewbridge⁴, Timothy D. Leeds⁴ and Kyle E. Martin⁵, (1)USDA-ARS-NEA-NCCCWA, KEARNEYSVILLE, WV, (2)USDA Agricultural Research Service, Kearneysville, WV, (3)USDA-ARS-NCCCWA, Kearneysville, WV, (4)ARS-USDA, Kearneysville, WV, (5)Troutlodge, Inc., Sumner, WA

Egg quality traits including survival rate to eyeing stage, fecundity and egg size are important to rainbow trout breeding companies, but present challenges for genetic analyses because they are measured only on sexually-mature fish and the number of phenotyped fish is usually limited. This study used pedigrees and phenotypes (N=7,905 records) from multiple generations in eight cohorts (4 strains × 2 year classes) from Troutlodge, Inc to estimate heritabilities and genetic correlations. In addition, ~100 phenotyped fish from each cohort were genotyped (57K SNP chip) to elucidate genetic architecture using multiple regression single-step genome-wide association analyses. Heritability estimates were 0.27±0.04, 0.49±0.04, and 0.51±0.04 for eyeing rate, fecundity and egg size, respectively. Significant negative genetic correlation (-0.34±0.06) was observed between fecundity and egg size with dam spawning weight having a significant positive genetic correlation with fecundity (0.31±0.08). Weak positive genetic correlations were observed between eyeing rate and either fecundity or egg size (0.19±0.08 for both). The genetic architecture of the traits was polygenic with some moderate-large effects QTL. One large-effect fecundity QTL with effective genetic variance (EGV) of 26% was identified on Omy05 within the double-inversion region that covers most of the chromosome. The strongest-effect eyeing rate QTL was on Omy26 (EGV=14.5%) and on Omy25 (EGV=9.7%) and Omy28 (EGV=8.2%) for egg size. Our results indicate that progress can be accomplished through pedigree-based selective breeding for those egg quality traits. In addition, the approach of combining genotyping and phenotyping data across-populations for genomic selection models may improve accuracy of genetic merit estimates.

PE0434: Aquaculture

Comparative Transcriptomic Analysis in Rainbow Trout Skeletal Muscle Reveals Opposing Functions Mediated By Glucocorticoid and Mineralocorticoid Receptors

Jorge Eduardo Aedo, Alfredo Molina and **Juan Antonio Valdés**, Universidad Andres Bello, Facultad de Ciencias de la Vida. Interdisciplinary Center for Aquaculture Research (INCAR), Santiago, Chile

Cortisol is an essential hormone and has a wide variety of effects on most tissues in vertebrates. Cortisol exerts its effects through the activation of glucocorticoid receptor (GR) and mineralocorticoid receptor (MR), however the role of both receptors in skeletal muscle is unknown. To understand in a comprehensive and global manner how GR and MR modulates the skeletal muscle transcriptomic response, we performed an RNA-seq analysis. Juvenile rainbow trout (*Oncorhynchus mykiss*) were intraperitoneally injected with physiological doses of cortisol (1 mg/kg). We also include a pre-treatment with mifepristone (GR antagonist) and spironolacton (MR antagonist). Samples were obtained 12 hours after injection. cDNA libraries were constructed from the skeletal muscle of rainbow trout groups: vehicle, cortisol, spironolactone, mifepristone, mifepristone/cortisol and spironolacton/cortisol. In total, HiSeq sequencing generated 280,334,668 paired-end reads. These reads were preliminary analyzed with the CLC Genomic Workbench software using a reference transcriptome for rainbow trout skeletal muscle resulting in ~86.1% of the reads mapped. The expression level of each transcript was represented as RPKM, and the analysis revealed that 1271 and 979 transcripts were differentially expressed mediated by GR and MR, respectively. In GR group, BP were significantly enriched in ubiquitin-dependent protein catabolic process (GO:0006511), myofibril assembly (GO:0030239), autophagy (GO:0006914). In MR group, BP were significantly enriched in mitotic nuclear division (GO:0007067), nuclear division (GO:0000280), muscle system process (GO:0003012), striated muscle cell development (GO:0055002). These results suggest as a whole that GR and MR have a differential participation in fish muscle growth. This study was funded by CONICYT, FONDAP 15110027 and FONDECYT 1171318

PO0435: Aquaculture

RNA-Seq of Bacterial Outer Membrane Vesicles and Whole Transcriptome Identify sRNAs Targeting Immune Relevant Genes in Rainbow Trout Susceptible or Resistant to *Flavobacterium psychrophilum*

Pratima Chapagain, Middle Tennessee State University, Murfreesboro, TN, Ali Ali, Animal and Avian Sciences, College Park, MD and Mohamed Salem, University of Maryland, College Park, MD

The outer membrane vesicles (OMVs) of gram-negative bacteria contain sRNAs, toxins, and virulence factors that are released from the bacterial surface during host-pathogen interaction. In this study, OMV from *Flavobacterium psychrophilum* (Fp), the etiological agent of BCWD in salmonids, was isolated and visualized by Transmission Electron Microscopy (TEM) as small spherical Nano-shaped particle around 50-100nm. RNA-Seq identified 750 transcripts expressed in OMVs and 1,940 transcripts expressed in transcriptome with 700 transcripts in common. Three transcripts associated with OMVs formation (OmpA family protein, ompA family outer membrane protein) have been highly expressed in the transcriptome.

Out of 236 sRNAs computational identified in the Fp genome, 5 sRNAs were expressed in the OMVs, and 10 sRNAs were expressed in the transcriptome.

The potential interaction of sRNAs-trout immune genes was investigated in two rainbow trout genetic lines, BCWD-resistant and susceptible, created by selective breeding. qPCR was used to determine the reciprocal expression of the sRNAs and their mRNA targets in whole-body lysates after 5 days of infection.

Interestingly, immune-related genes, mitogen-activated protein kinase-7 (MAPK), and Macrophage mannose receptor-1, c-c motif chemokine 21-like were downregulated in the susceptible compared to resistant fish whereas sRNAs (expressed in OMVs) targeting those genes were showing the reciprocal expression. Further research is suggested to explain the mechanism of sRNAs entry through vesicles into host cells.

PE0436: Aquaculture

Identification of Additional Single Nucleotide Polymorphisms Associated with Resistance to Bacterial Cold Water Disease in Rainbow Trout Using Whole Genome Resequencing

Sixin Liu¹, Guangtu Gao¹, Roseanna Long¹, Jason P. Evenhuis¹, Gregory D. Wiens¹, Kyle E. Martin² and Yniv Palti¹, (1)USDA-ARS-NCCCWA, Kearneysville, WV, (2)Troutlodge, Inc., Sumner, WA

Bacterial cold water disease (BCWD), caused by *Flavobacterium psychrophilum*, is a major disease in rainbow trout (*Oncorhynchus mykiss*). Previously, we have reported two major QTL associated with BCWD resistance on chromosomes Omy8 and Omy25. The objectives of this study were to identify additional single nucleotide polymorphisms (SNPs) associated with resistance to BCWD using whole genome resequencing, and to identify candidate genes for BCWD resistance. We conducted two rounds of pool-seq analysis in the Troutlodge odd-year May spawning population. For the first round of pool-seq, we pooled the DNA of parents based on their QTL haplotypes and the BCWD survival phenotypes of their offspring. In the second round of pool-seq, we pooled parental DNA samples solely based on BCWD phenotypes to avoid bias due to haplotype pre-selection. Over 10 million SNPs were identified in each round of pool-seq. Based on the first round of pool-seq, new SNPs showing significantly different allele frequencies between the two pools were used to genotype the 2015 Troutlodge May spawning population, and 26 SNPs associated with the BCWD resistance were validated. Candidate genes for the Omy08 QTL have also been identified after examining the functional annotation of the validated SNPs. Based on the second round of pool-seq, additional SNPs potentially associated with the two BCWD QTL have been identified, and we are currently evaluating those SNPs using the 2015 Troutlodge May spawning population.

PO0437: Aquaculture

Transcriptomic Response of Rainbow Trout Skeletal Muscle Challenged with *Piscirickettsia salmonis* Under Stress.

Alfredo Molina, Juan Antonio Valdés, Philip Dettleff and Rodrigo Zuloaga, Universidad Andres Bello, Facultad de Ciencias de la Vida. Interdisciplinary Center for Aquaculture Research (INCAR), Santiago, Chile

Stress and diseases are two of the main factors that affect Chilean aquaculture, generating important economic losses. Under stressful conditions, the immune response of the fish is affected, favoring the appearing of bacterial diseases such as Piscirickettsiosis, generated by the intracellular bacteria *Piscirickettsia salmonis*. Despite that skeletal muscle is the main product in aquaculture, little information exist respect on how this tissue response against the infection with this pathogen and how an increase in cortisol levels (main hormone associated with stress response) affects the response on myotubes during the infection with *P. salmonis*. To answer this interrogates, we develop an *in vitro* experiment using Rainbow trout myotubes. Four experimental groups were used as follow: (i) untreated cell or control (CTRL); (ii) *P. salmonis* ATCC LF-89 (MOI 50)-infected cells for 8 hours (INF); (iii) 100 ng/ml cortisol-treated cells for 3 hours (CORT) and (iv) pre-treated cells with cortisol and then infected by *P. salmonis* (CORT+INF). After treatments, we collected RNA samples, performed RNA-seq analysis with Illumina Hiseq, and then analyzed by CLC Genomic Workbench platform. The RNA-seq analysis showed several differentials expressed genes among the groups. A higher transcriptional response was observed across the groups associated with cortisol pre-treatment. Interestingly, we observed that cortisol impacts on the transcriptional capacity of myotube to response to *P. salmonis*, modulating processes associated to immune response and apoptosis. This results evidence that myotube-pathogen interaction could be modulated by cortisol, information that should be considered to understand the effect of stress in bacterial infections in muscle of salmonids species. Funding: CONICYT FONDECYT 1171307, CONICYT FONDECYT 1171318 and CONICYT/FONDAP 15110027.

PE0438: Aquaculture

Insights into Sex Determination in the Siamese Fighting Fish *Betta splendens*

Young Kwon¹, Carla R. Hoge¹, Hannes Svoldal², Richard Durbin³ and Andres Bendesky⁴, (1)Columbia University, New York City, NY, (2)University of Antwerp, Antwerp, Belgium, (3)University of Cambridge, Cambridge, United Kingdom, (4)Columbia University, New York, NY

Sex determination in teleosts is labile and diverse, including several types of monofactorial to polygenic systems that can differ within a genus and even species. The Siamese fighting fish *Betta splendens* have been domesticated for hundreds of years leading to morphological and behavioral features that are highly sexually dimorphic relative to their wild counterparts. The genetic determinants that drive sex differences is currently unknown in *Betta*. Genome wide analysis of whole genome sequences of 18 male and 15 female ornamental *Betta splendens* reveal two genomic

loci associated with sex. Through a genotyping screen of an additional 100 unrelated individuals and comparative phylogenomics, we examine the penetrance and stability of sex-specific haplotypes across ornamental and wild *Betta splendens* and the *Betta* phylogeny. We are currently developing genetic tools to assess the functional significance of candidate genes located in these haplotypes. Through the investigation of the genetics underlying sex determination and their evolution in *Betta splendens*, we hope to provide insight into the mechanisms that drive the diversity of sex determination systems in fish.

PO0439: Aquaculture

Investigating the Genetics of *Betta splendens* Red/Blue Coloration

Carla R. Hoge¹, Young Kwon¹, Hannes Svoldal², Richard Durbin³ and Andres Bendesky⁴, (1)Columbia University, New York City, NY, (2)University of Antwerp, Antwerp, Belgium, (3)University of Cambridge, Cambridge, United Kingdom, (4)Columbia University, New York, NY

The domestication of the Siamese fighting fish, *Betta splendens*, has been characterized by artificial selection for a variety of traits, including color, fin morphology, and aggressive phenotypes. In particular, just over 100 years of selection for ornamental features in *Betta* fish has resulted in dramatic variation in coloration, including hue, saturation, brightness, iridescence, and patterning. Whereas classical genetics has described the pattern of inheritance of some coloration phenotypes, nothing is known about the genomic loci controlling these traits in *Betta*. Using whole genome re-sequencing of 33 independent ornamental *B. splendens* at >15x coverage, we investigated the genetic components that drive the differences in color between red and blue fish. We called variants relative a wild *B. splendens* reference genome. After filtering this dataset, we located 4 areas of the genome associated with red/blue coloration across three different methods: Genome-wide F_{st}, GWAS, and KMER analysis. Our current work uses biochemical and genetic tools to validate the functional significance of candidate genes within these peaks. This study will not only serve as a proof of principle to demonstrate our ability to identify and perturb genes associated with phenotypic variation in *B. splendens*, but also will characterize morphological variation that has arisen during a recent domestication event.

PE0440: Aquaculture

Mechanisms of Benthic Adaptation and Frequent Molting in Penaeid Shrimps

Jianbo Yuan, Institute of Oceanology, Chinese Academy of Sciences, Qingdao, China and Prof. Jianhai Xiang and Fuhua Li's Group

Crustacea, the subphylum of [Arthropoda](#) which dominates the aquatic environment, is of major importance in ecology and fisheries. Here we report on the genome sequence of the Pacific white shrimp *Litopenaeus vannamei*, covering ~1.66 Gb (scaffold N50 605.56 Kb) with 25,596 protein-coding genes and the highest proportion of simple sequence repeats (>23.93%) among sequenced animals. The noted expansion of genes related to vision and locomotion is probably central for its benthic adaptation. Frequent molting of the shrimp is associated with an intensified ecdysone signal pathway through gene expansion and positive selection. As an important aquaculture organism, *L. vannamei* has been subjected to high selection pressure during the past 30 years of breeding, and this has had a significant impact on its genome. Decoding the *L. vannamei* genome not only provides an insight into the genetic underpinnings of specific biological processes, but also provides valuable information for enhancing crustacean aquaculture.

PO0441: Aquaculture

A Fast Sex Detection Method for the Whiteleg Shrimp *Litopenaeus vannamei* By Post-PCR High Resolution Melting (HRM)

Ricardo Perez-Enriquez¹, Raul Llera-Herrera², Pavel Galindo Torres¹ and Susana Avila¹, (1)Centro de Investigaciones Biológicas del Noroeste, S.C., La Paz, BS, Mexico, (2)Instituto de Ciencias del Mar y Limnología-UNAM, Mazatlan, SI, Mexico

The Pacific Whiteleg shrimp *Litopenaeus vannamei* is one of the most important aquaculture species globally, for which monosex female culture is desirable due to its higher weight at harvest. The development of technologies for

the production of sex-biased populations requires a verification protocol based on molecular sex markers, particularly during early life stages when no secondary sexual characteristics are displayed (i.e. petasma in males). In the present study, a 542 bp putatively sex-linked locus was sequenced for eight males and four females. The composite genotype of three single nucleotide polymorphisms (SNP) followed the ZZ/ZW sex-determination system (males were homogametic and females heterogametic). A post-PCR high resolution melt (HRM) analysis was designed to amplify and differentiate homozygotes from heterozygotes of the SNP in the locus. The expected HRM patterns for homozygotes and heterozygotes was confirmed in male and female wild and cultivated shrimp samples (n = 168). A blast search of the 542 bp sequence did not match to previously described sex-associated linkage groups. These results are of interest for sex detection and for future detailed studies on the sex-determination system in *L. vannamei*.

PE0442: Aquaculture

Development of Efficient Tools for the Molecular Breeding in *Litopenaeus vannamei*

Yang Yu, Quanchao Wang, Qian Zhang, Zheng Luo, Jianbo Yuan, Xiaojun Zhang, Fuhua Li and Jianhai Xiang, Institute of Oceanology, Chinese Academy of Sciences, Qingdao, China

The whiteleg shrimp *Litopenaeus vannamei* (*L.vannamei*) is a predominant aquaculture shrimp species in worldwide. It is considered as the aquaculture species with the highest single output value. Advances in selective breeding have accelerate the development of *L. vannamei* aquaculture. The molecular breeding method such as Marker Assisted Breeding (MAS) and the Genomic Selection (GS) method has been widely used in plant and animal breeding which have accelerated the breeding accuracy and efficiency. However, only a few researches on molecular breeding were reported in shrimp. In the past years, we have developed series of tools for the molecular breeding of *L.vannamei*. We developed an efficient method for family and broodstock identification based on the microsatellite markers, which was useful for the genetic assignment and relationship evaluation. Plenty of SNPs were discovered based on the genome re-sequencing data and a 600K high density SNP chip were designed. We also developed an efficient median density SNP genotyping method based on target sequencing approach. A lot of SNPs associated with economic traits could be genotyped simultaneously. Based on the high throughput genotyping methods, we have identified several growth related genes including *PKC*, *Rap-2a* and *SRC* and identified the SNPs in *ALF*, *TRAF6*, *PI3K* associated with disease resistance. The accurate MAS method for the growth and disease resistance traits selection were developed. We conducted the GS analysis and established the GS method in *L.vannamei*. The GS method has been proved to be a promising method for the genetic selection. Our research have solved several technique problems in molecular breeding of shrimp and the developed molecular tools will play important roles in the further genetic selection of *L.vannamei*.

PO0443: Aquaculture

DNA Methylation Dynamics in Atlantic Salmon (*Salmo salar*) after Being Challenged with High Temperature and Moderate Hypoxia

Anne Beemelmans¹, L. Ribas², D. Anastasiadi³, J. Moraleda-Prados², F.S. Zanuzzo¹, M.L. Rise¹ and A.K. Gamperl¹, (1)Department of Ocean Sciences, Memorial University of Newfoundland, Canada, (2)Institut de Ciències del Mar, ICM-CSIC, Spain, (3)The New Zealand Institute for Plant and Food Research, New Zealand

The marine environment is predicted to become warmer and more hypoxic over this century, and these conditions may become a challenge for cultured Atlantic salmon by negatively affecting their growth, immunology and welfare. DNA methylation mediates phenotypically plastic responses in gene expression that can potentially facilitate acclimatization responses. Thus, we measured DNA methylation from salmon that were subjected to: i) control conditions (normoxia, 12°C); ii) an incremental increase in temperature (12°C to 20°C, at 1°C per week) and then held at 20°C for 4 weeks; or iii) the former temperature regimen in combination with moderate hypoxia (~70% air saturation). DNA methylation levels were measured at CpG sites within a ~500 bp region (Promotor, 5'UTR, Exon, Intron) of six important liver biomarker genes (*cribp*, *jund*, *pkd3*, *prdx6*, *serpinh1*, and *ucp2*). Considering both experimental groups, we found 12 CpGs (out of 94 total) across the six genes that were differentially methylated when exposed to 20°C for 3 days, whereas only 6 CpGs from three genes (*jund*, *prdx6* and *ucp2*) were affected after 4 weeks at 20°C. At both time points, we uncovered distinct DNA methylation profiles for fish of each

treatment group, suggesting that high temperature and moderate hypoxia were inducing different CpG methylation changes in the liver of salmon. Further, we report significant relationships between CpG methylation and the mRNA expression of these genes that are complex and dynamic. These changes in DNA methylation may be an important regulatory mechanism allowing Atlantic salmon to quickly respond to new environmental challenges associated with global warming.

PE0444: Aquaculture

Genome of the Soft-Shell Clam and Its Transmissible Cancer

Samuel FM Hart, University of Washington, SEATTLE, WA

Cancer is normally an evolutionary dead-end—neoplastic cells that arise and evolve within an organism either regress or kill their host, and the death of the host marks the death of the cancer lineage. However, in some cases, neoplastic cells develop the ability to spread from individual to individual, turning from conventional cancers into clonal contagious cancer lineages. The natural transmission of cancer cells has been observed in two mammals (Tasmanian devils and dogs), and we have found that leukemia-like diseases in at least five bivalve species are due to the horizontal spread of clonal cancer lineages. One lineage affects soft-shell clams (*Mya arenaria*) along the east coast of North America and is ultimately fatal in most clams, contributing to the depletion of this commercially harvested species in many areas. We are currently assembling a reference genome for the soft-shell clam using PacBio sequencing combined with HiC data. Using draft reference genomes, we are investigating genomic changes in the evolution of this unique cancer lineage, including SNPs, structural variation, and copy number variation. We have found a retrotransposon, *Steamer*, which is expressed and amplified in genomic DNA of the contagious cancer lineage, expanding from 2-10 copies per haploid genome in normal animals to >100 in neoplastic cells. These new integration events and other genomic changes have likely played a role in oncogenesis and continued evolution of the cancer with its hosts.

PO0445: Aquaculture

Development of an Easy Genotyping Method for a Candidate Gene Responsible for Benedenia Disease Resistance in the Japanese Yellowtail, *Seriola quinqueradiata*.

Ryo Okada, Tokyo university marine science and technology, Tokyo, Japan

Gene selection directly selects individuals with a target economically valuable trait using the gene and gene variants responsible for the target trait as DNA markers. When selection and breeding programs are carried out using genotyping of a gene responsible for a target trait, the effects of inbreeding can be avoided. Marine fish of the genus *Seriola*, commonly known as yellowtail or amberjacks, are one of the most important species in aquaculture. Benedenia disease is a serious problem for the aquaculture of yellowtail/amberjacks and is caused by infection by the

monogenean parasite *Benedenia seriolae*. Previous quantitative trait locus (QTL) mapping of resistance traits for Benedenia disease revealed that the QTL with the largest genetic effect is located in the proximal region of the Squ2 linkage group in Japanese yellowtail. Detailed analysis of the genomic region of the QTL strongly suggested that the C-type lectin gene controls resistance to Benedenia disease. A single nucleotide polymorphism in the C-type lectin domain is significantly associated with *B. seriolae* infection level. In the N-terminal of the C-type lectin gene there is a

repetitive element composed of five amino acids, and the function of this domain is unknown. The aim of the present

study is to develop gene selection markers to establish yellowtail strains that retain resistance to Benedenia disease.

PCR-based easy and low-cost genotyping methods for these variants of the C-type lectin gene were developed.

Aquaculture-scale genotyping in the wild Japanese yellowtail population under phenotypic selection for Benedenia disease resistance suggests that the repetitive element tends to be associated with parasite infection levels. Thus, our genotyping methods are applicable for large-scale screening of yellowtail from the wild population, and gene selection

to establish a strain with the Benedenia disease resistance trait from a wild population using Japanese yellowtail and closely related species.

PE0446: Aquaculture

An Integrated Analysis of mRNA, lncRNA, and miRNA Expression Profiles in *Edwardsiella ictaluri*-Resistant and -Susceptible Catfish Identified By Marker Assistant Selection

Wenwen Wang, Auburn University; The Fish Molecular Genetics and Biotechnology Laboratory, Auburn, AL

Enteric septicemia of catfish (ESC), caused by bacterial pathogen *Edwardsiella ictaluri*, is one of the most serious diseases for the catfish industry. Ictalurid catfish are the primary aquaculture species in the US. Genome-wide association study analyses were previously conducted with channel catfish, first and third generation of backcross catfish progenies in our lab and significant QTL/SNPs associated with ESC resistance were identified on linkage group (LG) 1, facilitating marker-assisted selection (MAS) as a solution for genetic enhancement of ESC resistance. In this study, we distinguished resistant and susceptible catfish individuals through MAS at 0 h and 4 h after infection with *E. ictaluri*, and mRNA, lncRNA and miRNA expression profiles were analyzed in liver and intestine using high-throughput technologies. In liver, a total of 85 and 21 genes, 96 and 80 lncRNAs and 15 and 16 miRNAs showed significant differential expression between susceptible and resistant fish after 0 h and 24 h infection, respectively. In intestine, a total of 588 and 78 differentially expressed genes, 146 and 125 differentially expressed lncRNAs and 5 and 0 differentially expressed miRNAs were identified after 0 h and 24 h infection, respectively. GO and KEGG biological pathway analysis were performed to predict the functions of differentially expressed lncRNAs and co-expressed potential targeting genes. Co-expression networks of lncRNA-mRNA and miRNA-mRNA were constructed based on the correlation analysis between the differentially expressed RNAs. The present study provided an insight into the function of coding and non-coding RNAs between resistant fish and susceptible fish.

PO0447: Aquaculture

Chromosome-Level Assembly of *Periophthalmus magnuspinnatus*: An Indigenous Mudskipper in the Yellow Sea

Chul Lee¹, Juwan Kim¹, Byung June Ko², Heebal Kim³, Erich D. Jarvis⁴ and Vertebrate Genome Project, (1)Interdisciplinary Program in Bioinformatics, Seoul National University, Seoul, South Korea, (2)Department of Agricultural Biotechnology and Research Institute of Agriculture and Life Sciences, Seoul National University, Seoul, South Korea, (3)Seoul National University, Seoul, South Korea, (4)The Rockefeller University, New York, NY

Giant-fin mudskipper, *Periophthalmus magnuspinnatus* (PM), is an important marine fish. It lives endemically in coasts of the Yellow Sea, adapts to both in and out of water, and has a potential as a bio-indicator to monitor environmental changes. The previous reference genome of PM provided chances to understand molecular mechanisms of its land adaptation, but the short-read based assembly was too fragmented to analyse genomic contexts at chromosome level. In order to correct the mis-assembly, we generated a *De Novo* chromosomal scale genome assembly of PM (fPerMag1) by using 4 long read sequencing technologies. The 753 Mb genome met the VGP platinum quality (3.4.2.QV40 phased metric) with a contig N50, a scaffold N50, and a quality value of 2.3 Mb, 32.9 Mb, and 44.8, respectively. Based on Hi-C data, the number of chromosomes of PM was estimated as 25. The fragmented BUSCO genes were decreased as 2.4% from 4.5% in fPerMag1 compared to the old assembly. We believe fPerMag1 provides an unprecedented opportunity for cytogenomics approaches to investigate chromosomal evolution across Gobiidae fishes.

PE0448: Aquaculture

Genome Assembly, Genome-Wide SNP Identification and Population Structure Analysis in Yellowtail (*Seriola quinqueradiata*)

Tsubasa Uchino¹, Masatoshi Nakamoto², Junko Tabata¹, Yukinori Shimada³, Kazunori Yoshida³, Koichi Okuzawa³, Toshiya Suzuki⁴, Tsutomu Noda⁵, Yuichiro Fujinami⁵, Takashi Sakamoto² and Akiyuki Ozaki¹, (1)National Research Institute of Aquaculture, Fisheries Research Agency, Mie, Japan, (2)Department of Marine Biosciences, Tokyo University of Marine Science and Technology, Tokyo, Japan, (3)National Research Institute of Aquaculture, Fisheries Research Agency, Oita, Japan,

(4)National Research Institute of Aquaculture, Fisheries Research Agency, Kagoshima, Japan, (5)Seikai National Fisheries Research Institute, Fisheries Research Agency, Nagasaki, Japan

The yellowtail (*Seriola quinqueradiata*) is one of the most important aquaculture fish in Japan. One of main traits of interest assessed in the genetic breeding program in yellowtail is rapid growth by QTL for marker-assisted selection, or genomic prediction and genomic selection. Furthermore, the development and utilization of single nucleotide polymorphism (SNP) array enables the increased genotype accuracy of genome-wide SNP markers. Construction of the chromosome-scale genome assembly and identification of high-density SNP markers are necessary steps to design SNP-array of yellowtail.

We constructed *de novo* assembly of yellowtail by the whole genome shotgun method using Illumina short read and Pacbio long read sequences. In addition, we newly constructed genetic linkage maps using three full-sib families and the genotyping-by-sequencing method. The scaffold ordering and orientation of the scaffolds in the newly constructed genetic linkage maps yielded a total genome of 659.7 Mbp, with 2,603 scaffolds, and an N50 scaffold size of 27.9 Mbp.

We carried out *de novo* SNP discovery of yellowtail with whole genome resequencing with Illumina short reads using 106 unrelated fish samples from Pacific coastal areas of Japan. These unrelated fish become founder population in *Seriola* genetic breeding program in our institutes. We identified a total of 3.6 million SNPs in the yellowtail genome assembly and evaluated genome-wide linkage disequilibrium (LD) patterns. Population genetic structure analysis showed unclear genetic differentiation among several wild subpopulations of yellowtail. The genomic and SNP resources created in this study will be useful for further genetic improvement of yellowtail.

PO0449: Poultry

A Multiomics Toolbox to Advance Avian Research

Ashley E Meyer¹, LaDeana Hillier¹, Erich D. Jarvis², Giulio Formenti², Olivier Fedrigo², Bettina Haase², Jacquelyn Mountcastle², Jennifer Balacco², Christine G Elsik³, Yvonne Drechsler⁴, Nicholas Anthony⁵, Ron Okimoto⁵, Arang Rhie⁶, Adam M. Phillippy⁶, Hans H. Cheng⁷ and Wesley Warren¹, (1)Animal Sciences, Bond Life Sciences Center, University of Missouri, Columbia, MO, (2)Vertebrate Genome Laboratory, The Rockefeller University, New York, NY, (3)Division of Animal Sciences, University of Missouri, Columbia, MO, (4)College of Veterinary Medicine, Western University of Health Sciences, Pomona, CA, (5)Cobb-Vantress, Siloam Springs, AR, (6)Genome Informatics Section, NHGRI, NIH, Bethesda, MD, (7)USDA-ARS Avian Disease and Oncology, East Lansing, MI

Recent technological advancements led us to develop new genomic resources to significantly enhance avian genomics. First, to reduce sequence gaps and missing microchromosome assignments that exist in the GRCg6a reference, we constructed single-haplotype genome references for a commercial layer (paternal) and broiler (maternal) line. Measures of sequence contiguity completeness show N50 contig 'ungaped' lengths of 16 and 14 Mb in paternal and maternal genomes, respectively. After iterative scaffold builds with BioNano and HiC data, we achieve the near theoretical chromosome N50 scaffolds lengths of 60 and 89 Mb for paternal and maternal genomes, respectively. Second, using single cell RNAseq (scRNAseq) technology to reconstruct intimate details of resident cell types, we are building a deeper understanding of the cell type specific molecular circuitry of the avian immune system. Two adult chicken tissue types were initially investigated, liver and spleen, to optimize and compare single cell and nuclei methods. We find approximately 9 (primarily hepatocytes, 72%) and 10 (primarily erythroid-like, 69%) different cell types in liver and spleen, respectively. Cell identities were mostly inferred from comparisons to human and mouse, but will require further experimental validation. Finally, without easily accessible data portals, these resources cannot meet the needs of avian researchers. We have started to integrate chicken genomic resources into the established InterMine database, a multispecies domestic animal genome browser with extensive data mining capabilities. These chicken resources will collectively empower biological studies to fully

understand the avian cellular circuitry that emboldens the health readiness of the poultry industry.

PE0450: Poultry

Genome-Wide Identification and Annotation of Functional Regulatory Regions in the Chicken, Cattle, and Pig Genomes

Colin Kern¹, Ying Wang², Xiaoqin Xu², Zhangyuan Pan², Michelle M. Halstead³, Ganrea Chanthavixay², Perot Saelao², Susan Margaret Waters⁴, Nares Trakooljul⁵, Klaus Wimmers⁵, Ian Korf⁶, Mary E. Delany², Hans H. Cheng⁷, Juan F. Medrano³, Alison Van Eenennaam⁸, Christopher K. Tuggle⁹, Catherine W. Ernst¹⁰, Pablo J. Ross³ and Huaijun Zhou³, (1)Animal Science, University of California Davis, Davis, CA, (2)Animal Science, University of California, Davis, CA, (3)Department of Animal Science, University of California, Davis, CA, (4)University of California, Davis, Davis, CA, (5)Leibniz-Institute for Farm Animal Biology (FBN), Dummerstorf, Germany, (6)The Genome Center, University of California, Davis, Davis, CA, (7)USDA, ARS, Avian Disease & Oncology Laboratory, East Lansing, MI, (8)University of California, Davis, CA, (9)Animal Science, Iowa State University, Ames, IA, (10)Department of Animal Science, Michigan State University, East Lansing, MI

Only a small fraction of animal genomes is known to be transcribed, however the non-transcribed regions play key roles in gene regulation, and therefore significantly impact phenotypic traits. Epigenetic marks are a primary factor in identifying these regulatory regions, and while consortia such as ENCODE have made great progress in generating epigenetic data in human, mouse, and other model organisms, very little has been done in farm animals where researchers would benefit greatly from such data by improving the understanding of economically important traits. The Functional Annotation of ANimal Genomes (FAANG) consortium was created to coordinate epigenomics research in domesticated animals. As one of the FAANG pilot projects, we have completed the first comprehensive identification of regulatory elements in farm animals which included eight assays profiling the transcriptome, four histone modifications, CTCF, DNA methylation (RRBS) and open chromatin across eight tissues in the chicken, cattle, and pig genomes. Integration of these data produced genome-wide chromatin state predictions resulting in catalogs of promoters, enhancers, insulators, and polycomb repressed regions for each tissue in each species. These data and results are publicly available on the FAANG data portal and viewable on genome browsers via a UCSC track hub. The computational pipeline has also been made available on GitHub for use by the research community. These datasets, which are being expanded by ongoing FAANG-related projects, provide a resource for further studies that will improve the understanding of complex traits and the evolution of regulatory elements.

PO0451: Poultry

Combined Omic Analysis Characterizes Hepatic Maturation in the Post-Hatch Broiler Chick

Heidi A Van Every¹, Allen H. Hubbard², Justine Tarsillo³, Ashley Taylor³ and Carl J. Schmidt⁴, (1)Center for Bioinformatics and Computational Biology, University of Delaware, Newark, DE, (2)Donald Danforth Plant Science Center, Olivette, MO, (3)University of Delaware, Newark, DE, (4)Dept. of Animal & Food Sciences, University of Delaware, Newark, DE

High-throughput omic technologies have vastly expanded the knowledge gained from biological studies. Although individual types of omic data offer specialized insights, integration of these data types can provide a systems-level representation of biological processes and expedite hypothesis generation. In the modern broiler chicken, selected over decades for high feed efficiency and muscle growth, we employ transcriptomics and metabolomics to contrast hepatic metabolism and organ development at two time points post-hatch. At Day 4, lipid metabolites derived from the yolk remnant are stored in the liver to fuel organ growth while the bird's digestive system is maturing. The transcriptome is enriched for cell proliferation supporting organ growth processes. At Day 20, the liver displays a metabolic profile more typical of adult function and carbohydrate metabolism, and increased immune-related gene expression. We also apply a pipeline for unsupervised statistical analysis of metabolite ratios in the context of biological pathways to detect compounds involved in differential fates. This combined, data-driven approach prioritizes both genes and metabolites while locating them within experimentally testable biological relationships.

PE0452: Poultry

Functional Analysis Indicates SAP30 As Putative Causal Gene for Muscle Growth in Chicken

Bruna Petry¹, Aline G L Copola², Gabriel C. M. Moreira¹, Thais F Godoy¹, Luan G Clemente¹, Erika Cristina Jorge², Jane de Oliveira Peixoto³, Monica Correa Ledur³, James E. Koltes⁴ and Luiz L. Coutinho¹, (1)University of São Paulo, Piracicaba, Brazil, (2)University of Minas Gerais, Belo Horizonte, Brazil, (3)Embrapa Swine and Poultry, Concórdia, Santa Catarina, Brazil, (4)Iowa State University, Ames, IA

Chicken is an important source of high-quality protein and muscle growth is a central trait for poultry breeding programs. The goal of this study was to validate the role of the *SAP30* gene on muscle cell growth and differentiation. Integrative analysis of genomic studies previously performed in a Brazilian broiler population indicated *SAP30* as a positional candidate gene for breast muscle weight and percentage. In addition, this gene has a biological process associated with skeletal muscle cell differentiation. Gene editing was performed in mouse immortalized myoblasts (C2C12), using CRISPR/Cas9 to modify *SAP30*. The edited cells for *SAP30* and control group were transfected using Lipofectamine 3000 and the cells were incubated in 5% CO₂ at 37°C for 3 days in differentiation medium (DMEM, with 2% of horse serum). Three mutations were generated by gene editing and confirmed by Sanger DNA sequencing. Immunocytochemistry and morphometric analyses were conducted on MF20-positive myotubes and fusion index and myotube area analyses were performed. Statistical analysis comparing myotube area size between the control and treated groups were performed using ANOVA and Dunnett's test. An increase in myotube area was observed ($p < 0.05$) in the *SAP30* knockout clones. Moreover, we determined that in at least one of the clones, the mutation altered predicted protein structure. Our results indicate that *SAP30* plays a role in muscle cell growth affecting myotube area and likely contributes to the hypertrophy phenotype in chickens.

PO0453: Poultry

Analysis of Chicken Sperm Transcriptome By Super-Seq

Yunlei Li, Yanyan Sun, Aixin Ni, Hui Ma and **Jilan Chen**, Institute of Animal Science, CAAS, Beijing, China

It has been evidenced in several species that a mature sperm contains small quantity but complex RNA populations. The transcriptome of spermatozoa was believed to have a close relationship with semen quality and fertility, and the RNA could function as a carrier of epigenetic information. This study was conducted to characterize the transcriptome of the chicken mature sperm with Single-cell universal poly(A)-independent RNA sequencing (SUPeR-seq) method. Semen samples were collected from four Beijing-You cocks by abdominal massage. After washed by 5 mL PBS for two times, the samples were purified with the somatic cell lysis buffer. The spermatozoal RNA was isolated using TRIzol reagent. The results indicated that the chicken spermatozoal RNA lacks 18S and 28S ribosomal RNA, and RIN value is 2.5. The majority of the transcripts ranged between 25 to 500 nt. RNA libraries were prepared and sequenced on Illumina platform to obtain 150 bp pair end reads. Each sample produced 34 to 69 M clean reads, 18% of which were mapped to the gallus genome (*Gallus gallus* GRCg6a) with tophat2 software. A total of 9,300 mRNA transcripts were detected. The transcripts shared by at least three samples with FPKM >10 were selected for further analysis. GO enrichment analysis showed that these transcripts were associated with intracellular signal transduction, actin cytoskeleton organization, cilium morphogenesis and assembly, egg activation, ciliary basal body, cilium, ubiquitin protein ligase activity and ubiquitin conjugating enzyme binding. KEGG pathway analysis indicated that these transcripts were involved in glycerophospholipid metabolism, mTOR signaling pathway, and ubiquitin mediated proteolysis. A total of 700 known and 1,000 novel lncRNA transcripts were detected. The target genes of high-expressed lncRNAs (FPKM >10) were enriched in nucleus, cytosol, cilium, actin cap, metal ion binding, nucleotide binding, regulation of phosphoprotein phosphatase activity and maintenance of protein location in nucleus GO terms. This study presents for the first time the mRNA and lncRNA transcripts of chicken spermatozoa with SUPeR-seq technology and suggests the potential association between sperm transcripts and spermatogenesis.

PE0454: Poultry

Genetic Basis of Wooden Breast and White Striping in Commercial Broilers

Juniper A. Lake¹, M. Walugembe², Luke Kramer², Jack C.M. Dekkers³ and Behnam Abasht⁴, (1)Center for Bioinformatics and Computational Biology, University of Delaware, Newark, DE, (2)Department of Animal Science, Iowa State University, Ames, IA, (3)Iowa State University, Ames, IA, (4)University of Delaware, Newark, DE

Wooden breast (WB) and white striping (WS), the two major myopathies of fast-growing broilers, continue to threaten global poultry production due to their severe impacts on meat quality. Tightly associated with economic traits such as growth rate, feed efficiency, and breast muscle yield, these muscle disorders present an exceptional challenge to producers, as dietary or management strategies against WB and WS generally impair performance. An understanding of the genetic basis of these myopathies is therefore critical for developing a long-term solution. To this end, a genome-wide association study was conducted using 1,193 Cobb500 broilers, which were raised to seven weeks of age and evaluated for WB, WS, and body weight. Genotyping was performed using restriction enzyme-based genotyping by sequencing and quantitative trait loci (QTL) were detected using single-SNP analysis and Bayesian multi-marker regression (BayesB). Heritability was estimated to be moderate for all traits: 0.51 ± 0.06 for WB, 0.53 ± 0.06 for WS, and 0.41 ± 0.06 for body weight. Genetic correlation between WB and WS was high (0.87 ± 0.02), but genetic correlation between either WB or WS and body weight was low (0.24 ± 0.03 and 0.18 ± 0.03 , respectively). Multi-marker analysis of WB found eight 1-Mb regions that each explained $>1\%$ of genetic variance. Together, these eight regions explained 18.5% of the genetic variance for WB. Multi-marker analysis of WS and body weight identified, respectively, five and four 1-Mb regions that each explained $>1\%$ of genetic variance. These results improve our understanding of the genetic architecture of these myopathies and suggest that WB and WS would respond to selection.

PO0455: Poultry

***MC1R* Variants and Their Effects on Plumage Pigmentation in Native Japanese Chicken Breeds**

Md Humayun Kabir¹, Atsushi Takenouchi², Mohammad Ibrahim Haqani¹, Yoshiaki Nakamura^{1,2}, Sakae Takeuchi³ and Masaoki Tsudzuki^{1,2}, (1)Graduate School of Biosphere Science, Hiroshima University, Higashi-Hiroshima, Hiroshima, Japan, (2)Graduate School of Integrated Sciences for Life, Hiroshima University, Higashi-Hiroshima, Hiroshima, Japan, (3)Graduate School of Natural Science and Technology, Okayama University, Okayama, Japan

Native Japanese chicken breeds are unique in plumage coloration. Melanocortin 1 Receptor (*MC1R*) is the key gene to regulate melanin synthesis. In total, 304 and 92 chickens from 12 Japanese- and 5 non-Japanese breeds, respectively, were used to investigate the *MC1R* gene sequence and their association with feather color. Three synonymous (C69T, G636A & C834T) and 10 non-synonymous (G178A, T212C, G274A, G376A, T398A, G409A, A427G, T637C, A644C & C919G) nucleotide substitutions resulted in 14 haplotypes (H0 – H13), based on the haplotypes of Red Junglefowl (H0 & H1). Wild-type plumage patterned Japanese breeds Tosa-Jidori and Onagadori possessed mostly H0, while-Ryujin-Jidori was exceptional (H1). However, Brown Leghorn possessed both. Gifu-Jidori exhibits wild-type and wheaten plumage, consequently, possessed H0 and H2 (A427G, G636A & T637C). The wheaten and brown colored breeds, Ko-Shamo, Nagoya, and Kumamoto mostly exhibited H2 as like Rhode Island Red, while-Mie-Jidori showed H4 (A644C). Among black plumaged breeds, Tômaru presented birchen allele (H3: T398A, G636A, T637C & C834T)) as like Black Minorca. Yakido was notable for new H8 (T212C, G274A & G636A with SNP T637). The Kurekadori had mostly H6 (C69T, T212C & G274A) with H7 (C69T, T212C, G274A, G636A & T637C) representative of extended black, though, the plumage is black-breasted-silver. Contrarily, Australorp was fixed with H6. The white Ukokkei possessed novel substitution C919G, causing two new haplotypes (H12: C69T, G178A, G274A, G636A, T637C & C919G; H13: C69T, G274A, G409A & C919G). In conclusion, Japanese chicken breeds showed three new haplotypes (H8, H12 & H13), with various unique features.

PE0456: Poultry

Molecular Analysis of the CPQ Gene, a Quantitative Locus for Pulmonary Hypertension in Broilers

Duaa Almansaf, UNIVERSITY OF ARKANSAS BIOLOGICAL SCIENCES DEPT, Fayetteville, AR and Shatovisha Dey, Alia Parveen, and Douglas Rhoads

Abstract

We have further analyzed the expression of CPQ gene to understand the basis for this locus contribution to ascites phenotype in broilers. Ascites syndromes is a pathophysiological condition results in the mortality rate of fast-growing broilers. Ascites results from increasing the body demand for oxygen and the accumulation of fluid in the abdominal cavity after increasing the pressure within the pulmonary circulation. In our lab, Whole Genome Resequencing (WGR) identifies genetic and chromosomal regions related to ascites. We found that there are 31 regions associated with ascites in our unselected relax line (REL). One of these regions, is around 127 Mbp on chromosome 2, where more than 40 SNPs show both ascites resistance and susceptible in males. Most of these SNPs are in the downstream region of Carboxypeptidase Q gene (CPQ) or Plasma Glutamate Carboxypeptidase gene (PGCP), and this gene belongs to the family of M29 peptidase. CPQ gene encodes a metallopeptidase which plays a significant role in the cleavage of dipeptides into amino acids. Also, this gene is responsible for releasing thyroxine from thyroglobulin. In our work, we extracted DNAs from 12 REL line birds in both genders and sequenced the DNA in order to have tissues samples from heterozygous and homozygous. Then, we collected eight tissues from each bird including heart, liver, kidney, thigh, breast, spleen, lung, and thymus to see the expression level of CPQ gene in all eight tissues through using SNPs in intron 6 and exon 8 of this gene. We compared the CT values of this gene with the housekeeping gene. Our results indicated that there is a normal expression of the CPQ gene in all eight tissues even though a previous study in our lab showed there are a high expression of the CPQ gene in lung, heart, and liver. Therefore, we decided to use RNAseq analysis because we can get additional information beside the gene expression including mutations and allele specific expression.

PO0457: Poultry

Identification of Active Promoters and Enhancers By H3K27ac Peaks in the Spleen Tissue of Two Inbred Chicken Lines Under NDV Infection and Heat Stress

Ying Wang¹, Colin Kern², Karen Tracy¹, Ganrea Chanthavixay¹, Perot Saelao¹, Susan J. Lamont³, R.a. Gallardo⁴, Nathan Chubb⁵, Gonzalo Rincon⁵, T.R. Kelly⁴ and Huaijun Zhou², (1)Animal Science, University of California, Davis, CA, (2)Department of Animal Science, University of California, Davis, CA, (3)Department of Animal Science, Iowa State University, Ames, IA, (4)University of California, Davis, School of Veterinary Medicine, Population Health and Reproduction, Davis, CA, (5)Zoetis, Kalamazoo, MI

Newcastle disease (ND) and heat stress are two critical factors that decrease poultry meat and egg production. The objective of this study was to identify genome-wide H3K27ac histone modifications in the spleen to discover changes in promoter and enhancer activity in response to NDV infection under heat stress using two highly inbred chicken lines (Leghorn (susceptible) and Fayoumi (resistant)). At 21 days, both were inoculated with NDV under heat stress. ChIP-seq for H3K27ac was performed on spleens collected at 6 days post infection (6DPI) for two individuals from each treatment-line group (4). Peaks were called using the MACS2 peak caller (q value < 0.01) for each sample. Genes with active promoters were identified by a H3K27ac peak within 2kb of the transcription start site (TSS), while genes with active enhancers had a H3K27ac peak between 2kb to 100kb of the TSS which was not near to another gene's TSS. Preliminary analysis showed there were 11,156 genes with active promoters and 13,465 genes with active enhancers. Over 80% of genes with active promoters, but only 20% of genes with active enhancers, were consistently identified in the 4 groups. 1,016 genes had an active promoter and enhancer across all 4 groups. Gene ontology analysis showed that these genes are mostly involved in host immune functions, including T cell receptor and Toll-like receptor signaling pathways. Further investigation on the line-specific regulatory elements changes due to treatment will help elucidate gene regulation associated with resistance or susceptibility to the treatment in chicken spleen.

PE0458: Poultry

Investigation of Gut Integrity and Cellular Tight Junctions in Stressed Broiler Chickens

Amer Hasan^{1,2}, Douglas D. Rhoads³, Nnamdi Simon Ekesi³ and Adnan Alrubaye⁴, (1)University of Arkansas, Fayetteville, AR, (2)University of Baghdad. Veterinary Medicine college. Department of Veterinary Public Health, Baghdad, Iraq, (3)University of Arkansas, Cell and Molecular Biology program, Fayetteville, AR, (4)University of Arkansas/Fayetteville, Fayetteville, AR

We have been investigating integrity of gut epithelia of broiler chickens grown on wire and litter flooring. Previous work has demonstrated that growth on wire flooring dramatically increases the incidence of bacterial chondronecrosis with osteomyelitis (BCO). Our hypothesis is that wire flooring induces stress that allows bacteria to cross epithelia into the bloodstream, and colonize the growth plates of the rapidly growing leg bones. The integrity of the epithelial cell layers that protect the body from the external microorganism is preserved by intercellular junctional complexes (tight junctions, adherens junctions). The goal of this study is to assess the status of tight junction and adherens proteins for broilers raised on different flooring. The results will help us understand how stress can induce BCO lameness. Lameness is one of the main metabolic diseases linked to fast growth in broilers. It is a significant problem in the poultry industry, resulting in hundreds of millions of dollars in lost revenue annually. My research is using immunohistochemistry and immunofluorescence to assess gut integrity and quantify the levels of tight junction and adherens proteins. We have administered different probiotics, and prebiotics to broiler birds on both floor types, with some positive reduction in BCO lameness. The investigation of tight junctions under different conditions will help us understand how these treatments reduce BCO lameness.

PO0459: Poultry

Chromosome-Level Genome and Comparative Analyses of Duck Uncover Genome and Chromatin Architecture Evolution of Birds

Jing Li¹, Jilin Zhang², Zhengkui Zhou³ and Qi Zhou¹, (1)Life Sciences Institute, Zhejiang University, Hangzhou, China, (2)Department of Medical Biochemistry and Biophysics, Karolinska Institutet, Solna, Sweden, (3)Institute of Animal Science, Chinese Academy of Agricultural Sciences, Beijing, China

Duck is one of the most important poultry species and also a key model to study avian influenza viruses, yet its current genome quality is much lower than those of most birds. Here we combined diverse technologies (PacBio, 10X, BioNano and Hi-C) and RH linkage map, and produced a high-quality chromosome-level genome for Peking duck. The new genome (ZJU1.0) has been improved for contiguity by 70-fold greater than the previously published version (BGI1.0) and dramatically refines the gene annotation. We found that the specific burst of certain subfamilies (CR1 and ERVL) of repetitive elements in duck is the major cause of fragmentation of the old genome. We also identified the putative telomere and centromere regions, and specific CR1 family sequences that are enriched at the centromeres of microchromosomes. Comparison of chromatin architectures between chicken and duck revealed strong natural selection against the inversions that disrupt the topologically associated domains. Sex chromosome analysis indicated that the duck W chromosome is not fully degenerated. In total, we identified 84.5 Mb Z chromosome which including 2.2 Mb pseudoautosomal regions (PAR) and 16.8 Mb W chromosome. We demarcated three times of recombination suppression based on the Z/W sequence similarity, forming a pattern of ‘evolutionary strata’ along the duck Z chromosome. Overall, the new duck genome provides an important resource to study avian genome evolution and future improvement of domesticated traits.

PE0460: Poultry

Antimicrobial Activities of Duck Liver-Expressed Antimicrobial Peptide 2 (LEAP-2) Against Pathogenic Bacteria

Yeong Ho Hong¹, Yeojin Hong¹, Anh D. Truong¹ and Hyun S. Lillehoj², (1)Chung-Ang University, Anseong, South Korea, (2)USDA-ARS, Beltsville, MD

Antimicrobial peptides (AMPs) are important peptides of innate immune systems. AMPs are known to have antimicrobial activity against Gram-negative bacteria, Gram-positive bacteria, fungi, viruses, parasites, and even cancer cells. We identified and characterized the duck LEAP-2 gene to determine its antimicrobial activity against Gram-positive and Gram-negative bacteria. Tissue samples were collected from 6–8-week-old Pekin duck (*Anas platyrhynchos domesticus*), and total RNA was extracted to cDNA synthesis. Quantitative real-time-PCR was conducted to confirm the duck LEAP-2 transcript expression levels. Two kinds of peptides (a linear peptide and a disulfide-type peptide) were synthesized and antimicrobial activity assay and fluorescence microscopic analysis were conducted to demonstrate duck LEAP-2 bactericidal activity. The duck LEAP-2 peptide sequence showed high identity with those of other avian species over 85%, as well as more than 55% of identity with mammalian sequences. LEAP-2 mRNA was highly expressed in the liver with duodenum next, and then followed by lung, spleen, bursa and jejunum. Both LEAP-2 antimicrobial peptides efficiently killed pathogenic bacteria, although the

disulfide-type LEAP-2 showed more powerful bactericidal activity. Also, Gram-positive bacteria was more susceptible to duck LEAP-2 than Gram-negative bacteria. Finally we confirmed that LEAP-2 peptides could kill bacteria by disrupting the bacterial cell envelope using microscopy. Taken together, duck LEAP-2 showed its antimicrobial activity against both Gram-positive and Gram-negative bacteria. In particular, disulfide bonds were important for a powerful killing effect by disrupting the bacterial cell envelope. Therefore, duck LEAP-2 can be used for effective antibiotics alternatives.

PO0461: Poultry

Directed Genome Evolution Identifies Deoxyribose Phosphate Aldolase As a Macrophage Survival Factor in *Staphylococcus agnetis*

Abdulkarim Shwani¹, Sura Zaki¹, Sonali Lenaduwe¹, Sohita Ojha² and Douglas D. Rhoads¹,
(1)University of Arkansas, Cell and Molecular Biology program, Fayetteville, AR, (2)Biology Department, Langston University, Langston, OK

Abstract

We have employed Directed Genome Evolution (DGE) to identify two copies of the deoxyribose-phosphate aldolase gene as essential factors for survival of *Staphylococcus agnetis* (isolate 908) in a chicken macrophage cell line. A specific amino acid substitution appears to be associated with survival and killing. Macrophage survival may be a key component of the hypervirulence of isolate 908 in inducing bacterial chondronecrosis with osteomyelitis (BCO) in broilers. We first reported the isolation of this staphylococcal species from the bones and blood of lame broilers at the University of Arkansas. We have demonstrated high incidences (>60%) of BCO through administration of *S. agnetis* 908 through aerosols or drinking water. The annotated complete genome of isolate 908 has been published. BCO primarily affects the growth plate of the proximal femur and tibia of fast-growing broilers, but survival in the blood may be essential for transmission to the growth plate. Our phylogenomic analyses of chicken and cattle isolates of *S. agnetis* and *Staphylococcus hyicus* suggest a very close relationship between cattle and chicken isolates. Chicken isolate, 908, is closely related to a cattle isolate, strain 1379. Yet, more than 40 genes and 3 plasmids from strain 908 are absent or poorly conserved in any of the cattle *S. agnetis* isolates. We have found that isolate 1379 is efficiently killed by chicken macrophage but isolate 908 not only survives phagocytosis, this isolate efficiently kills immortalized chicken macrophage within 2 days. We produced and sequenced more than 13 independent DGE transformants of 1379 with 908 DNA. Our analyses demonstrate that multiple elements can be efficiently transferred but the unifying property of all DGE selections is a single amino acid change in either copy of deoxyribose-phosphate aldolase. Our current efforts are aimed at confirmation using purified PCR products, and to understand the relevance of this “stress-response” determinant in survival and killing.

PE0462: Poultry

Defining QTLs, Genes and Genome Variants for Resistance to Marek's Disease Virus

Jacqueline Smith, The Roslin Institute, The University of Edinburgh, Edinburgh, United Kingdom, Ehud Lipkin, Hebrew University of Jerusalem, Ganigar, Israel, Morris Soller, Hebrew University of Jerusalem, Jerusalem, Israel, Janet E. Fulton, Hy-line International, Des Moines, IA and David W. Burt, University of Queensland, Brisbane, QLD, Australia

Marek's Disease Virus (MDV) is a highly-contagious oncogenic alpha-herpes virus, which infects chickens, causing neurological symptoms and tumour formation. It is also a highly immune-suppressive virus. Although Marek's disease (MD) is partially controlled by vaccination, it continues to have a large impact on animal health and a profound effect on the poultry industry in terms of profit and animal welfare. Even after years of study, the genetic mechanisms underlying resistance to MDV remain poorly understood. The MHC is known to play a role in disease resistance, and a handful of other genes have also been implicated in resilience. Genome wide association study (GWAS) of an F₆ advanced intercross line derived from a cross between two commercial layer lines phenotyped for survival in the face of MDV challenge has allowed us to define many Quantitative Trait Locus Regions (QTLR) associated with resistance. Within these QTLRs have identified candidate genes for the survival phenotype by integrating transcriptomics and genetic association tests. The Fms-Related Tyrosine Kinase 3 (*FLT3*) gene has been shown to have strong genetic association with resistance, along with several other candidate genes including both

miRNAs and lncRNAs. This provides targets for mitigating the effects of MDV on both animal health and the poultry economy by way of gene assisted breeding, improved vaccine design or gene-editing technologies.

PO0463: Poultry

Genetic Characterisation of Avian Leukosis Virus (ALV)-like Tumours Identified in Commercial Chicken Lines

Katarzyna Miedzinska-Bielecka¹, Richard Kuo², Janet E. Fulton³ and Jacqueline Smith², (1)The Roslin Institute, EH25 9RG, United Kingdom, (2)The Roslin Institute, The University of Edinburgh, Edinburgh, United Kingdom, (3)Hy-Line International, Dallas Center, IA

Avian Leukosis Virus (ALV) is an Alpha-retrovirus and is classified into two groups: exogenous (subgroups A-D, J and K) which are transmitted as infectious virus particles and endogenous (subgroup E) which is a retrovirus-like element within the genome and is genetically transmitted but not always expressed. The exogenous virus is capable of infecting a wide variety of avian species. Infection with ALV causes oncogenic transformation and immune-suppression, making infected animals more susceptible to other pathogens. However, even in the absence of exogenous ALV infections, spontaneous ALV-like tumours have been reported in commercial layer and broiler chickens as well as turkey flocks after vaccination. Understanding these ALV-like tumours formations could have major implications for both the poultry industry and for pharmaceutical companies who rely on egg production for vaccine development.

To understand genetic susceptibility to ALV-like tumour development, we analysed whole genome sequencing (WGS) from vaccinated commercial chickens. We compared 4 sample groups: blood from p27 negative ELISA tested chickens (p27-), blood from p27 positive ELISA tested chickens (p27+), ALV-like liver tumours (LT), and blood from the chickens with ALV-like liver tumours (C-). We identified several genomic regions which may be associated with ALV-like tumorigenesis.

In one region, we found variation between the tumour genotypes (LT) and the non-tumour genotypes (p27+, p27-, C-) within the ELOVL2 and ETV1 genes which are associated with carcinogenesis. This preliminary data will aid further investigations into the nature of these tumours and the genetic basis of ALV/ALV-like infections.

PE0464: Poultry

Genomic and Virulence Comparisons of Different Bacterial Isolates from BCO Lesions in Broilers

Nnamdi Simon Ekesi, University of Arkansas, Cell and Molecular Biology program, Fayetteville, AR

We used embryo lethality assay (ELA) to examine the virulence of bacteria isolates from bacterial chondronecrosis with osteomyelitis (BCO) lesions, the leading cause of lameness in broiler chickens. Lameness poses serious animal health and welfare issues, as well as, significant economic losses. Our hypothesis is that bacteria cross epithelia, and some survive in the blood to colonize the proximal growth plates through weaknesses in the vascular of the rapidly growing leg bones. We compared *Escherichia coli*, and *Staphylococcus* species, when we induced lameness using the wire-flooring model, as well as lameness outbreaks at commercial broiler farms. Differences in ELA, especially among the *E. coli* strains, prompted us to examine phylogenies using whole genome comparisons. *E. coli* isolates 1409 and 1413, from neighboring farms of the same integrator, showed very different ELA results and affiliate with divergent *E. coli* clades. *E. coli* isolate 1527, from a different farm and integrator, had similar ELA results to and a genome very similar to *E. coli* 1413. Isolate *Staphylococcus aureus* 1516 represented a common BCO isolate from the third farm and is mildly virulent in the ELA relative to *E. coli* 1413, *E. coli* 1527 and a very pathogenic human *S. aureus* isolate. The genome of *S. aureus* 1516 is most similar to isolates from deep wounds/lesions from chickens in Poland and more distantly to chicken isolates derived from human *S. aureus* in the United Kingdom. ELA allows virulence comparisons of distinct isolates when containment facilities are not available for live bird work.

PO0465: Poultry

Further Evaluation of Differences in Mitochondrial Biogenesis in Broiler Muscles that Correlate with Ascites Susceptibility

Layla Al-Mitib, University of Arkansas, Fayetteville, AR

The overall goal of this project is to correlate differences in mtDNA copy number we have observed with quantification of active mitochondrial enzymes. We are focused on breast, thigh, heart, and liver (as a control) in our REL line males. Previously we had reported that there are gender, tissue, and ontological differences in mtDNA copy number in our SUS and RES line birds. The SUS is susceptible to ascites, while the RES is resistant, the REL is the unselected source population. Ascites is a metabolic disease that is influenced by increased growth of muscles (meat) in broiler chickens. Also in ascites, many genes have reported high expression, such as *PPARGC1A* that have shown significantly high ratio in mtDNA/nucDNA in the breast muscle of 22 weeks ascites susceptible line males compared to resistant line males. Tissue samples have been dissected, stained and examined under microscope to investigate the mitochondrial quantities correlated to high number of mtDNA.

PE0466: Poultry

Genetic Diversity Study of Nigerian Turkey

Danlami Moses Ogah, Nasarawa State University Keffi, Animal Science Dept, Fac. of Agric. pmb 135 Lafia Nigeria, Lafia, Nigeria

MtDNA still represents a useful tool in the study of molecular genetic diversity, because it appears in multiple copies in the cells and the mitochondrial gene content is strongly conserved across generations. In order to understand the population diversity of Nigerian indigenous turkey breed and to preserve this genetic diversity, evaluation of mitochondrial DNA sequence was employed. The study was aimed at determining mitochondrial DNA (mtDNA) D-loop, HV1 region polymorphism of indigenous turkey population in Nasarawa State north central Nigeria. We analyzed the complete mitochondrial DNA D-loop. To achieve this, blood samples were collected from 30 indigenous turkey, 10 each from three different populations separated by distance. A 623-bp fragment of the mtDNA D-loop region was sequenced in the sampled turkey populations. The result obtained indicate that in total, 11 haplotypes were identified, Haplotype diversity and nucleotide diversity were 0.81 ± 0.07 and 0.15 ± 0.07 respectively. With 126 number of polymorphic sites. Analysis of Molecular Variance (AMOVA) based on partial D-loop sequences of the turkey population also indicates that 99.05% of the total sequence variation between haplotypes was present within the population and 95.00% between populations. These results show a high mitochondrial D-loop diversity and indicate multiple maternal origins for Nigeria indigenous turkey. The molecular information on genetic diversity revealed in this study may be useful in developing genetic improvement and conservation strategies to better utilize indigenous Nigerian turkey resources.

PO0467: Poultry

Quantitative Trait Loci for Growth Related-Traits in Japanese Quail (*Coturnix japonica*) Using Restriction-Site Associated DNA Sequencing

Mohammad Ibrahim Haqani¹, Md Humayun Kabir¹, Shigeru Nomura², Akira Ishikawa³, Michiharu Nakano², Atsushi J. Nagano⁴, Atsushi Takenouchi², Yoshiaki Nakamura^{1,2} and Masaoki Tsudzuki^{1,2}, (1)Graduate School of Biosphere Science, Hiroshima University, Higashi-Hiroshima, Hiroshima, Japan, (2)Graduate School of Integrated Sciences for Life, Hiroshima University, Higashi-Hiroshima, Hiroshima, Japan, (3)Graduate School of Bioagricultural Sciences, Nagoya University, Nagoya, Aichi, Japan, (4)Faculty of Agriculture, Ryukoku University, Otsu, Shiga, Japan

This study aimed to identify QTLs for growth-related traits by constructing a genetic linkage map based on single nucleotide polymorphism (SNP) markers in the Japanese quail. A QTL mapping population of 277 F₂ birds was obtained from an intercross between a male of a large-sized strain and three females of a normal-sized strain. Body weight was measured weekly from hatching to 16 weeks of age. Five non-linear regression growth models (Weibull, Logistic, Gompertz, Richards and Brody) were analyzed, and the Gompertz was selected as the best model for describing a quail growth curve of the F₂. Restriction-site associated DNA sequencing (RAD-seq) developed 125 SNP markers that were informative between two parental strains. Map Manager QTX b20 software constructed 16 linkage groups of the SNP markers that spanned 795.9 centiMorgan (cM) with an average marker interval of 7.3 cM. QTL analyses of 22 phenotypic traits (17 weekly-measured body weights and 5 Gompertz model parameters) using R/qtl package revealed 8 significant QTLs at genome-wide 5% level. Detected QTLs affected all traits except for

body weights at hatching and 1 week. These results can be helpful in identifying genes underlying the QTLs and in the application of marker-assisted selection in the Japanese quail.

PE0468: Poultry

A Premature Stop Codon of the *tr* Gene Drives Iris Color Diversity in the Rock Pigeon (*Columba livia*)

Si Si^{1,2}, Xiao Xu², Jian-Ou Gao³ and Shu-Jin Luo², (1)Academy for Advanced Interdisciplinary Studies, Peking University, Beijing, China, (2)The State Key Laboratory of Protein and Plant Gene Research, School of Life Sciences, Peking-Tsinghua Center for Life Sciences, Peking University, Beijing, China, (3)Runyujia Technology Co. Ltd, Beijing, China

Pigmentation color exhibits vast diversity in Columbidae and plays a vital role in selective breeding of the rock pigeon (*Columba livia*). There are three iris colors commonly observed in domestic pigeons, or, pteridine-based yellow, guanines-based pearl, and pigment-absent bull iris. Breeding records have established that yellow and pearl iris traits are determined by a single locus termed *tr* while the bull eye is by an independent locus epistatic to *tr*. However, the exact genetic basis underlying such variation has eluded us. Here we conducted genome-wide association study (GWAS) based on whole genome sequencing of domestic pigeons and identified the *tr* gene. Haplotype analysis showed a 1.1 Mb region strongly associated with recessive pearl eyes in pigeons (N=54) against the yellow-eyed wildtypes (N=52). Allele-specific differential expression validated that the nonsense-mediated mRNA decay (NMD) is caused by a premature stop codon in the *tr* gene, which might abolish the pteridines synthesis but exert no effect on accumulating guanines in iris chromatophores. The phylogeny based on a 21-kb nonrecombining *tr* region clustered the wildtypes and pearl-associated mutants into two distinct clades, indicative of a single origin of the trait. Furthermore, the low nucleotide diversity and strong extended haplotype homozygosity (EHH) of pearl-iris-associated haplotypes jointly suggested a partial selective sweep of the *tr* gene in domestic pigeons, which could be the result of artificial selection for pearl over the wild-type yellow eyes. This study revealed the genetic basis of the pearl-eyed *tr* gene and shed light on the selection-driven morphological diversity during animal domestication.

PO0469: Poultry

Genomic, and Virulence Comparisons of Different Bacterial Isolates from BCO Lesions in Broilers

Nnamdi Simon Ekesi, University of Arkansas, Cell and Molecular Biology program, Fayetteville, AR

We used embryo lethality assay (ELA) to examine the virulence of bacteria isolates from bacterial chondronecrosis with osteomyelitis (BCO) lesions, the leading cause of lameness in broiler chickens. Lameness poses serious animal health and welfare issues, as well as, significant economic losses. Our hypothesis is that bacteria cross epithelia, and some survive in the blood to colonize the proximal growth plates through weaknesses in the vascular of the rapidly growing leg bones. We compared *Escherichia coli*, and *Staphylococcus* species, when we induced lameness using the wire-flooring model, as well as lameness outbreaks at commercial broiler farms. Differences in ELA, especially among the *E. coli* strains, prompted us to examine phylogenies using whole genome comparisons. *E. coli* isolates 1409 and 1413, from neighboring farms of the same integrator, showed very different ELA results and affiliate with divergent *E. coli* clades. *E. coli* isolate 1527, from a different farm and integrator, had similar ELA results to and a genome very similar to *E. coli* 1413. Isolate *Staphylococcus aureus* 1516 represented a common BCO isolate from the third farm and is mildly virulent in the ELA relative to *E. coli* 1413, *E. coli* 1527 and a very pathogenic human *S. aureus* isolate. The genome of *S. aureus* 1516 is most similar to isolates from deep wounds/lesions from chickens in Poland and more distantly to chicken isolates derived from human *S. aureus* in the United Kingdom. ELA allows virulence comparisons of distinct isolates when containment facilities are not available for live bird work.

PE0470: Other Animal Species

The 2019 Illumina Agricultural Greater Good Initiative: An Overview

Ciani Elena, University of Bari "Aldo Moro", Bari, Italy, **Samantha A. Brooks**, University of Florida, Gainesville, FL, **Faisal Almathen**, King Faisal University, Al-Hasa, Saudi Arabia, **André Eggen**, Illumina, Paris, France and **Pamela A. Burger**, Research Institute of Wildlife Ecology, Vetmeduni, Vienna, Austria

Camels have served humans in cross-continental caravans, transporting people and goods, connecting different cultures and providing milk, meat, wool and draught since their domestication. Today, they still represent a key livestock resource in several low-income countries where, due to their peculiar biological and physiological features, specifically adapted to desert conditions, better than other livestock species can thrive and produce high-quality protein for food consumption. Moreover, rapid though fragmentary changes are ongoing in the camel sector, with a growing demand for sustainable milk and meat production. Therefore, camel farming systems in peri-urban areas are commonly evolving toward more intensive management practices. Implementation of Single Nucleotide Polymorphisms (SNPs) in a high-density genotyping platform may allow rapid and cost-effective genome-wide genotyping in large numbers of animals, thus boosting downstream applications such as genome-wide association studies for production traits and genome-based selective breeding. Most of the existing livestock SNP genotyping arrays have been generated through whole-genome re-sequencing efforts made by large international and well-established consortia. The camel scientific community is still in its infancy and has to face major structural and conjunctural issues that make access to research funding not an easy task. An unprecedented opportunity has been offered by the 2019 Illumina Agricultural Greater Good Initiative grant, that will allow whole-genome re-sequencing of more than 400 samples, representative of the entire geographic range of the camel distribution, for a total throughput of 20 tera-bases of Illumina NovaSeq sequencing data, as a first step toward the development of an Illumina® CamelHD BeadChip and the realization of the first genome-wide camel diversity study at such a large geographic scale. This study will also contribute to deepen the understanding of evolutionary processes that shaped the camel genomes and to decipher the molecular basis of the peculiar physiological adaptation traits of camels. Parallel progresses in large-scale recording of camel phenotypes following standardized procedures are necessary for an ecologically and economically sensible exploitation of the Illumina® CamelHD BeadChip genotyping tool.

PO0471: Other Animal Species

Functional Annotation of Non-Coding Genomic Regions of Non-Human Primates

Naoki Hirose¹, Masatsugu Ema², Erika Sasaki³, Oleg Gusev^{4,5}, Jun Kawai⁵, Yasuhiro Murakawa¹ and Hideya Kawaji^{1,6}, (1)RIKEN Center for Integrative Medical Sciences, Yokohama, Japan, (2)Shiga University of Medical Science, Shiga, Japan, (3)Central Institute for Experimental Animals, Kawasaki, Japan, (4)RIKEN-KFU Translational Genomics Unit, Yokohama, Japan, (5)RIKEN Preventive Medicine and Diagnosis Innovation Program, Yokohama, Japan, (6)Tokyo Metropolitan Institute of Medical Science, Tokyo, Japan

To understand functional non-protein-coding regions in genomes, the FANTOM5 Consortium has produced a series of genome-wide maps of human and mouse promoters, enhancers, long non-coding RNAs (lncRNAs) and microRNAs (miRNAs) expressed in various samples including the tissues (Forrest et al. 2014; Andersson et al. 2014; Arner et al. 2015; Hon et al. 2017; De Rie et al. 2017). CAGE (Cap Analysis of Gene Expression) was used to detect capped-5' ends of RNAs in a nucleotide resolution, in addition to RNA-seq and small RNA-seq. Despite such functional annotations in the both species, non-coding regions are less conserved in the human and the mouse genomes. Therefore, it has been obstructive to thoroughly explore evolutionary landscapes of the functional non-coding regions of mammalian genomes.

To change this situation, we are producing an atlas of functional genomics of non-human primates, crab-eating macaques (*Macaca fascicularis*) and common marmosets (*Callithrix jacchus*). We applied CAGE, RNA-seq and small RNA-seq on various tissues of the macaques and the marmosets, respectively, majority of which matched the human's tissues investigated by FANTOM5. Here we determine >100,000 promoters, >10,000 enhancers, >20,000 lncRNAs and >3,000 miRNAs of the macaques and the marmosets, respectively. The human promoters, enhancers, lncRNAs and miRNAs are conserved 41~90% in the macaque, 29~84% in the marmoset and 12~68% in the mouse, respectively. Our annotations on a genome of the non-human primates may deepen insights into evolution of the functional non-coding regions of mammalian genomes.

PE0472: Other Animal Species

Genomic Analyses between *Neofelis nebulosa* and *Neofelis diardi* Aid in Conservation Efforts for Clouded Leopards

Madeline Grace Bursell¹, Rebecca Dikow², Warren E. Johnson³, Klaus-Peter Koepfli⁴ and Paul B. Frandsen¹, (1)Brigham Young University, Provo, UT, (2)Smithsonian OCIO Data Science Lab, Washington, DC, (3)The Walter Reed Biosystematics Unit, Museum Support Center MRC-534, Smithsonian Institution, Suitland, MD, (4)Center for Species Survival, Smithsonian's National Zoological Park and Conservation Biology Institute, Washington, DC

Due to decreasing costs in genome sequencing, conservation genomics is a field that has experienced immense growth over the last few years. By comparing whole genome data within and among threatened and endangered populations, we can estimate important elements in conservation such as levels of heterozygosity and demographic histories that reveal the level of endangerment of a species. Looking at the distribution of heterozygosity across a genome can help to reveal the levels of inbreeding within a population. This information informs conservation priorities and captive breeding programs. In the present study, we focus on two species of clouded leopards: *Neofelis nebulosa* and *Neofelis diardi*. *N. nebulosa* is a species of clouded leopard that lives in mainland southeast Asia. *N. diardi* inhabits the islands of Sumatra and Borneo in Indonesia. While these two species were initially thought to be a single species, evidence, such as differences in fur color and sizes of cloud markings, hint that they have diverged into two species. To shed more light on the genomic differences between them, we sequenced, assembled, and annotated whole genomes from both species. With genomes on hand, we explore differences in demographic histories, variation in blocks of homozygosity, and generate a whole genome phylogeny with other large cat species. Using these analyses, we share insights that inform the conservation status of the two species.

PO0473: Other Animal Species

miRNAs Targeting the Hippo Signaling Pathway during Myoblast Differentiation Are Regulated By Changes of Chromatin States

Huanhuan Zhou, Yue Xiang, Mingyang Hu, Yueyuan Xu, Ye Hou, Xiaolong Qi, Yu Luan, Zhangxu Wang, Xinyun Li, Yunxia Zhao and Shuhong Zhao, Key Laboratory of Agricultural Animal Genetics, Breeding, and Reproduction of the Ministry of Education & Key Laboratory of Swine Genetics and Breeding of Ministry of Agriculture, Huazhong Agricultural University, Wuhan, China

miRNAs reportedly participate in various biological processes, such as skeletal-muscle proliferation and differentiation. However, the regulation of differentially expressed (DE) miRNAs and their function in myogenesis remain unclear. Herein, miRNA expression profiles and regulation during C2C12 differentiation were analyzed by RNA-seq, ATAC-seq, and CHIP-seq from the viewpoint of chromatin states. We identified 19 known and 9 novel DE miRNAs at days 0, 1, 2, and 4. Results showed that the expression of DE miRNAs was related to chromatin states of their surrounding 113 ATAC-seq peak-defined open chromatin regions. Among them, 44.25% were co-localized with MyoD/MyoG binding sites. The remaining part of the above open chromatin regions was detected to be enriched with motif of myoblast-expressed AP-1 family, CCCTC-binding factor (CTCF), CCCTC-binding factor like (CTCFL), and Bach2 transcription factors (TFs). The target genes of the above DE miRNAs were also primarily enriched in muscle growth and development pathways, especially the Hippo signaling pathway. This pathway was primarily downregulated according to the mRNA-sequencing dataset, in contrast to the primarily upregulated DE miRNAs involved in this pathway. Further, we showed that the upregulation of *Wnt1* mediated the increased expression of *MyoD* and *MyoG*. Our study provided new insights into the DE miRNA function and their expression alteration by chromatin states and TF binding during myoblast differentiation.

PE0474: Other Animal Species

Phenotype and Genome Analysis of Dwarfism in Alpacas

Michael Mee¹, Craig Sergeant¹, T. Y. Kevin Tan¹ and **Kylie Munyard²**, (1)Curtin University, School of Pharmacy & Biomedical Sciences, Curtin Health Innovation & Research Institute, Perth, WA, Australia, (2)Curtin University, School of Pharmacy & Biomedical Sciences, CHIRI Bioscience, Perth, WA, Australia

Approximately 50% of Herdsire X's cria are noticeably smaller than normal. Both male and female cria are affected, and there is also perceived mild dysmorphia. We report here the investigation of this trait, including phenotypic characterization, inheritance pattern, and a preliminary genome wide association study. The animals in the study were Sire X, cria out of normal females that were sired by Sire X, cria out of these same dams by normal sires, and related and unrelated unaffected controls. Phenotype was characterized by measuring body proportions at age >1yo. Mode of inheritance was hypothesised from pedigree records. Whole genome sequencing was performed on 34 alpacas, including four confirmed dwarfs. SNPs were generated from reads mapped to the alpaca VicPac3.1 genome. A GWAS using PLINK is underway.

The affected alpacas conformed to a disproportionate dwarfism phenotype, namely a significant reduction in the length of the spine ($P=0.002$), the length of elbow to withers ($P=0.02$), spine to hind knee ($P=0.03$), and hock to hind fetlock ($P=0.01$). Pedigree analysis supported an autosomal dominant mode of inheritance.

PO0475: Other Animal Species

Development of a SNP Microarray for Alpacas

Marcos Calderon Montes¹, **Gustavo A. Gutierrez Reynoso**², Manuel J. More Montoya¹ and F. Abel Ponce de León³, (1)UNIVERSIDAD NACIONAL AGRARIA LA MOLINA, Peru, (2)UNIVERSIDAD NACIONAL AGRARIA LA MOLINA, LIMA, Peru, (3)Department of Animal Science, University of Minnesota, Saint Paul, MN

The production of alpaca fiber contributes to the economy of rural families in the highland areas of the Andes. The implementation of genetic improvement programs based on genomic selection could accelerate the genetic gain for fiber quantity and quality. The aim of this study was the discovery of single nucleotide polymorphisms (SNPs) and the development of a SNP microarray in alpacas. DNA samples from 150 white Huacaya alpacas originating from two Peruvian geographical Andean regions were obtained to generate ApeKI and PstI/MspI reduced representation libraries for each sample. Libraries were sequenced on a HiSeq 2500 utilizing v4 chemistry and 1x100 single end reads generating a mean read depth of ~6X per library. A bioinformatic analysis using a variant calling pipeline allowed to identify 4'283,956 variants across the VicPac3.1 alpaca reference genome (GCA_000164845.4). A list of 228,636 SNPs was generated considering the parameters phred-scaled quality score (>10), call rate (≥ 0.45), minor allele frequency (between 0.05 and 0.50), Illumina Design Score (≥ 0.6), and no other SNPs located within the 81 bp SNP sequence. Of these, 80,219 SNPs located at equidistant intervals of 25 Kbp were identified and will be part of the Alpaca SNPchip. In this manner, 97% of chromosomally assigned scaffolds and 78% of unassigned scaffolds are covered with SNPs, representing a genome coverage of 92%. Finally, we are also including SNPs localized in candidate genes for fiber growth and fiber color.

Keywords: alpaca, SNP, microarray

PE0476: Other Animal Species

An Overall Presentation of the T Cell Receptor (TR) Genomics in Camelidae: Future Perspectives through Illumina Sequencing in *Camelus dromedarius*

Giovanna Linguiti¹, Rachele Antonacci¹, Pamela A. Burger², Elena Ciani³, Serafina Massari⁴ and **Salvatrice Ciccicarese**¹, (1)Department of Biology, University of Bari, Bari, Italy, (2)Vetmeduni Vienna, Research Institute of Wildlife Ecology, Vienna, Austria, (3)Department of Biosciences, Biotechnologie and Biopharmaceutics, University of Bari, Bari, Italy, (4)Department of Biological and Environmental Science and Technologies, University of Salento, Lecce, Italy

Here we review the structural characteristics and the current status of the detailed genomic organization of α/β and γ/δ T cell receptors (TRs) loci in the *Camelus* genus. We discuss on the organization of loci and chains from γ/δ TR, during the synthesis of which somatic hypermutations occur, a process which is known for an extremely limited number of species for TR.

Thanks to identification of the reference (AMPH and STARD3NL) genes in *Camelus dromedarius* and *Camelus ferus*, we report the updating of the gamma (TRG) locus discussing evolutionary aspects with respect to human and sheep loci upon the identification of the TRGC5 cassette.

We also review the comparative genomics of the T cell receptor beta (TRB) locus starting from the *Camelus dromedarius* and moving forward through the draft genome sequences of its wild and domestic Bactrian congeners, *Camelus ferus* and *Camelus bactrianus*.

Therefore, with the aim of investigating the genomic variability of the T cell receptors loci in *Camelus dromedarius* considered as useful and promising model for therapeutic applications and for phylogenetic studies in the adaptive immune responses, we intend to take advantage of the large-scale project of Illumina genome sequencing. The 2019 Illumina® Agricultural Greater Good Initiative Award is expected to generate hundreds of whole-genome sequences of *Camelus dromedarius* samples collected from a wide intercontinental geographic range, thus paving the way toward an improved understanding of the genetic architecture of T cell receptors loci and their interindividual diversity.

PO0477: Other Animal Species

Efforts to Create Reference Quality Rat Genomes

Yanchao Pan¹, Bilge Ozel¹, Tristan Jong², David Ashbrook², Clifton Dalgard³, Leah Catherine Solberg Woods⁴, Hao Chen², Robert Williams², Abraham Palmer⁵, Huda Akil¹ and Jun Li¹, (1)University of Michigan, Ann Arbor, MI, (2)University of Tennessee Health Science Center, Memphis, TN, (3)Uniformed Services University, Bethesda, MD, (4)Wake Forest School of Medicine, Winston-Salem, NC, (5)University of California San Diego, La Jolla, CA

The common rat (*Rattus norvegicus*) is a key species in biomedical research. The rat reference genome (rn6) however has not been completed and its genome has ~100 times more fragments than either those of mouse and human. We performed whole-genome sequencing of the eight founders of the NIH Heterogeneous Stock. Heterozygote frequencies ranged from 10.6% (BN) to 15.0% (M520), much higher than expected for these inbred lines. Most heterozygous sites reside in 300+ intervals that are shared in almost all strains, that show higher read depths than average, and that contain higher rates of tri-allelic variants. These findings almost certainly result from mis-assemblies in rn6, specifically intervals in which two or more highly repetitive segments have been incorrectly combined. The known problems of rn6 motivated the establishment of an International Rat Omics Consortium (IROC) that is now using multiple technologies—PacBio, 10X Genomics Chromium, Bionano, HiC, Oxford Nanopore—to create much improved whole genome sequence for 80–100 inbred lines, and to bring them to reference quality. Our goal is to improve the genomic foundation for widely used inbred and outbred rat strains. With collaborators at Medical College of Wisconsin, UC Denver, the Czech Academy of Sciences and the University of Groningen, we have generated the first batch of 10X linked-read data for >30 lines. Variant calling and de novo assembly results are being analyzed. Analysis of structural variants is in progress. Also in progress is the hybrid assembly using 10X Genomics and BioNano data for select lines, to be augmented by PacBio, Nanopore and Hi-C data. Among strains in the queue are founders of HXB/BXH and LEXF families of strains, DSS/Mcwi and FHH/Mcwi lines, and the rest of the Hybrid Rat Diversity Panel (HRDP). These new data sets will require the refinement of workflows to integrate both diverse genomes and technologies, leading to the next generation of rat genome variant graph assemblies. In return, they will provide a greatly improved resources for research communities reliant on rat models.

(Supported by NIDA U01DA043098, U01DA047638, P30DA044223, and Center for Integrative and Translational Genomics at UTHSC)

PE0478: Other Animal Species

Whole-Genome Sequencing and Comparative Analysis of Emu Provides Insights into Sex Chromosome Evolution of the Palaeognathae Clade

Miki Okuno¹, Rei Kajitani¹, Hiroyuki Tanaka¹, Shuhei Mizushima², Asato Kuroiwa² and Takehiko Itoh¹,
(1)School of Life Science and Technology/Tokyo Institute of technology/Japan, Tokyo, Japan,
(2)Department of Biological Science/Hokkaido University/Japan, Hokkaido, Japan

Palaeognathous birds have morphologically conserved karyotypes and less differentiated ZW sex chromosomes. It has been reported that the sex chromosomes of ostrich and emu have exceptionally large recombining pseudoautosomal regions (PAR), while the rest of regions consist of strata 0 (S0) and strata 1 (S1). S0 corresponds to the most evolutionarily ancient stratum, while S1 referred to the evolutionarily younger stratum. However, the evolutionary path of the sex chromosomes in the Palaeognathae clade is unclear because the sequence diversity between the homologous chromosomes that makes sequence of the S1 region fragmented.

To understand sex chromosome evolution in the Palaeognathae clade, we constructed each of the Z/W sequences of the S1 region and performed comparative genome analysis. First, we carried out Illumina sequencing of emu genomes (male and female) and assembled the Z chromosome of emu (13 scaffolds, 82.4 Mbp in total) based on the whole genome sequencing data of the male individual. Next, the boundary between PAR and S1 was identified based on the copy number of the chromosome estimated using the sequence depth analysis of the female individual (ZW). In consequence, the boundary was found to be consistent in emu and cassowary, who belong to the same Casuariiformes order. In contrast, it was confirmed that the boundary of ostrich showed a relatively small PAR and an expanding S1 region. Furthermore, the Z/W sequence homology in the S1 region was found to be approximately 96%, 95%, and 89% for emu, cassowary, and ostrich, respectively.

PO0479: Other Animal Species

Genomics for the Conservation of Critically Endangered Great Indian Bustard

Yadvendradev Jhala¹, Harish Kothandaraman², Tripathi Abhishek², Bipin Kumar², Sutirtha Dutta¹, Yellapu Srinivas¹, Vishnu Priya Kolipakam¹ and **Dushyant Singh Baghel²**, (1)Wildlife Institute of India, India, (2)Nucleome Informatics Private Limited, Hyderabad, India

The Great Indian bustard (GIB) (*Ardeotis nigriceps*), Otididae family is endemic to India. Less than 150 individuals survive primarily in the western arid state of Rajasthan. The species is critically endangered and on Schedule I of the Wildlife Protection Act (1972). Primary threats to GIB are from power-line collisions, historical hunting, egg and chick predation by non-native predators, and habitat loss to agriculture. A conservation breeding program has been initiated as an insurance against extinction as well as to reintroduce/supplement wild populations after abating threats. Whole-Genome sequencing of one GIB using PacBio Sequel at 63X and Illumina at 82X using 10X genomics library and resequencing of two birds using Illumina Paired-End approach at 41x was done. The genome was assembled using WTDBG2, scaffolded with 10X datasets followed by polishing into 1269 scaffolds covering 99.16% of the genome having an N50 of 38Mb and a BUSCO score of 95% with the assembled genome size of 1.192 GB. Variant calling and tertiary analysis identified 3.3MB SNPs; 370,000 to 394,457 polymorphic SSR markers and 3000 SVs. Heterozygosity for the Great Indian Bustard at kmer 21 was 0.37% and for Houbara Bustard (*Clamydotos macqueenii*) was 0.56%. Using RNA Seq as transcript evidence, 16688 protein-coding genes were predicted and annotated. Similarly, the assembly of the mitochondrial genome and annotation resulted in the prediction of 14 protein-coding, 22 tRNAs and two rRNA genes.

Otididae seems to evolved 120 MYA during which time they have gained 317 gene families and lost 773 in comparison to eight available avian genomes that had 14924 gene families, with 4675 single copy ortholog families. GIB and Houbara appear to have diverged from a common ancestor 30 MYA. The GIB has 27 unique gene families, while Houbara has three unique gene families. PSMC revealed a population crash in GIB around 35000 YA which incidentally coincides with the advent of humans into India. Genomic data will assist in determining founders and pedigree to maintain heterozygosity in the captive population for successful conservation breeding program.

PE0480: Other Animal Species

Identification of Novel Functional SNPs in *Salamandra infraimmaculata* through RNA Sequencing

Valentina Rovelli¹, Vladimir V. Galatenko², Nadav Pezaro¹, Ori Segev¹, Arne W. Nolte³, **Tamar Krugman¹**, Alan Templeton¹ and Leon Blaustein¹, (1)Institute of Evolution, University of Haifa, Haifa,

Israel, (2)Lomonosov Moscow State University, Moscow, Russian Federation, (3)Carl von Ossietzky Universität Oldenburg, Institute for Biology and Environmental Sciences, Oldenburg, Germany

Whole Genome Sequencing and Reduced Representation Genome Sequencing are currently standard approaches in ecological genomics. Nevertheless, non-model species with large genomes still represent a challenge for the basic application of these methods. The Near Easter Fire Salamander (*Salamandra atra*) is an endangered amphibian with a genome size ~34 Gb, which requires massive computational analysis and high costs. With the aim of fostering new genomic studies on this species and identify novel functional SNPs, RNA sequencing and SeqSNP genotyping were applied to *S. atra* populations in Israel. RNA was sequenced from four pools of larvae (23, 50, 23 and 26 larvae respectively) sampled in the field, each pool representing a distinct geographic region in Northern Israel (Tel Dan, Upper Galilee, Lower Galilee, Mount Carmel). Putative SNPs were identified through the Kissplice pipeline and of those 500 were selected for the SeqSNP genotyping of 220 adult individuals from the four regions (14, 87, 43 and 76 respectively). After filtering for $maf < 0.05$, $DP > 15$, locus call rate $\geq 95\%$ and individual call rate $\geq 95\%$, a panel of 460 SNPs for 218 samples was obtained. Principal component analysis indicates that *S. atra* in Northern Israel is divided into four main genetic clusters. Tel Dan and Mount Carmel populations are isolated, while there is gene flow between the Upper and Lower Galilee populations, as evidenced by the proximity of their genetic clusters and previous analyses using microsatellites. Outlier analysis revealed the presence of 5 SNPs potentially under selection, which could shed light on local adaptation phenomena.

PO0481: Other Animal Species

Differential Gene Expression of Reverse Development in the 'Immortal Jellyfish' (*Turritopsis dohrnii*)

Yui Matsumoto and Maria Pia Miglietta, Texas A&M University, Galveston, Galveston, TX

The medusae (jellyfish) of *Turritopsis dohrnii* (Cnidaria, Hydrozoa) undergoes reverse development to avoid death caused by physical damage, adverse environmental conditions, or aging. Weakened or damaged jellyfish will undergo a whole-body transformation into a cluster of uncharacterized tissue, referred to as the cyst stage, which then will metamorphose back into an earlier lifecycle stage, the juvenile polyp. This unique ability has granted the species the name, the "Immortal Jellyfish". The underlying cellular mechanism that permits its reverse development is called transdifferentiation or cell reprogramming. Cell transdifferentiation allows fully mature and differentiated cells to reprogram themselves into a new cell type of any lineage. Thus, transdifferentiation is highly regarded in the biomedical sector as a potential mechanism to transform mature cells into any needed cell type after tissue damage. The polyp, medusa, cyst and reversed polyp stages of *T. dohrnii* were sequenced through RNA-sequencing and were assembled using a *de novo* approach.

Transcriptome profiling and time-series differential gene expression of *T. dohrnii*'s reverse development sequence revealed that the Cyst stage becomes enriched with transcripts that are associated with aging/lifespan, regulation of transposable elements, DNA repair and damage response and Ubiquitin-related processes, to name a few. Additionally, novel genes significantly upregulated during the rejuvenation process were identified in the Cyst. Ultimately, our work produced a high-quality reference transcriptome and foundation for *T. dohrnii* to develop an alternative model system to further investigate and understand regeneration, cellular plasticity and aging in metazoans.

PE0482: Other Animal Species

First Genome Release of Talitrid (*Trinorchestia longiramus*)

Ajit Kumar Patra¹, Oksung Chung², Ji Yong Yoo¹, Moon Geun Yoon¹, **Youngik Yang¹** and PAG Authors' Group, (1)National Marine Biodiversity Institute of Korea, Chungchungnam-do, South Korea, (2)Clinomics Inc, Ulsan, South Korea

Amphipoda is an order of malacostracan crustaceans usually found in aquatic habitats. Among amphipods, Talitridae is the only family exceptionally found in terrestrial and semi-terrestrial habitats and are often regarded as a vital group to study evolution. Talitrid species are considered as a key trophic link between primary producers and higher consumers, helping in the energy flow within the ecosystem. The genus *Trinorchestia* Bousfield, 1982 usually known as sand-hopper are mainly found in sandy beaches of South Korea and Japan. In this study, we

presents the first talitrid genome, the sand-hopper *Trinorchestia longiramus*. We generated ~380.3Gb of sequencing data and assembled to 0.89 Gb of the draft genome. By combining *ab initio* and homologous gene predictions 26,123 protein-coding genes were predicted from this genome. The completeness of the gene model was estimated with BUSCO, which resulted in 89.9%. Comparison with other amphipods showed that *T. longiramus* has 327 unique orthologous gene clusters, which are mostly expanded gene families responsible for the transport of toxicants from the cells, homeostatic process and tolerating ionic and osmotic stress. This first talitrid genome will be useful for further understanding the mechanisms of adaptation in the terrestrial environments as well as genomic variation from other amphipods.

PO0483: Other Animal Species

De Novo Transcriptome Analysis and Detection of Antimicrobial Peptides of the Red King Crab *Paralithodes camtschaticus*

Igor Yakovlev, Erik Lysøe, Snorre B. Hagen and Jihong Liu Clarke, Norwegian Institute of Bioeconomy Research, Ås, Norway

Endogenous antimicrobial peptides (AMPs) are evolutionarily ancient factors of innate immunity, which are produced by all multicellular organisms and play a key role in their protection against infection. Red king crab (*Paralithodes camtschaticus*), also called Kamchatka crab is a widely distributed and best known species of all king crabs belonging to the family *Lithodidae*. Despite their economic importance, only limited studies were reported. The genetic resources of king crabs are scarce and no full-genome sequences available to date. Therefore, identification and characterization of new AMPs from Red King crab transcriptome could potentially contribute to the future exploitation of AMPs from King crab and potential novel antimicrobial drug candidates when antibiotic resistance has become a global health threat.

In this study, we sequenced the *P. camtschaticus* transcriptomes from carapace, tail flap and legs tissues using an NGS platform. Libraries were systematically analyzed for gene expression profiles along with AMP prediction. To identify AMP candidates, we made blastp (protein-protein BLAST) search of AMPs from several databases against assembled contigs translated into proteins by all 6 ORFs. BLASTP revealed 417 contigs matching to 62 AMPs. Then we manually analyzed each of the AMP candidates and discarded all low similarity and redundant AMPs (i.e. different AMPs matching to the same crab contigs). In total, we defined 48 AMP candidates belonging to diverse families and functional classes, including: buforins, crustins, paralithocins, ALFs (anti-lipopolysaccharide factor), etc.

We analyzed expression patterns of 27 AMP genes. Of these, 19 genes were highly expressed, with more than 3000 reads in all libraries, 5 genes were expressed with around 1000 reads, and 3 genes had less than 100 reads and were lowly expressed. The highest expression was found for Paralithocin 1 and Crustin 3, with more than 8000 reads. Other Paralithocins, ALFs, Crustins and Ubiquicidins were among medium expressed genes. This transcriptome data set and AMPs provide a solid baseline for further functional analysis in *P. camtschaticus*. Results from the current study contribute also to the future application of King crab as a bio-resource in addition to be the known delicacy seafood.

PE0484: Other Animal Species

Unraveling the Genetic Basis of the Web-Building Behavior in the Spider *Uloborus Diversus*

Jeremiah D Miller, The Johns Hopkins University, Baltimore, MD and Andrew Gordus, The Johns Hopkins University

Our long-term goal is to establish the orb-weaving spider, *Uloborus diversus*, as a model organism for behavior, using web-building behavior to understand how the brain structures behaviors spanning multiple timescales. A significant challenge to studying such behavior in model organisms is that while they produce structured behaviors on short timescales (e.g., grooming, vocalizing, courting, etc.), these behaviors appear to be stochastically sampled on longer timescales, which introduces significant trial-to-trial variability. An ideal organism for studying long-timescale behaviors would produce a complex, reproducible, and quantifiable behavior over longer timescales. *U. diversus* is a simple organism which possesses these qualities. Web-building is a highly stereotyped behavior that

involves several carefully coordinated behaviors with defined steps, that are performed daily, and which produce a highly geometric structure: the web itself. Furthermore, previous research has shown that pharmacological perturbations of different neuromodulatory pathways have different and unique effects on the final structure of the web. While these studies demonstrated that psychoactive drugs dramatically alter specific aspects of web geometry, quantitative studies of the behaviors that produce these features are lacking. Using modern computer-aided vision, machine learning, and massively parallel computing tools, we are studying the web-building behavior itself in unprecedented detail. We hypothesize that perturbation of behaviorally relevant genes, using genetic tools such as RNA interference (RNAi) and CRISPR-Cas gene editing, especially neuromodulatory GPCRs and transporters, will produce detectable behavioral phenotypes that affect the same stages of web-building as previously conducted pharmacological studies. These phenotypes may provide insight into the genetic structure of innate, complex behaviors that are structured over multiple time scales. We would like to understand this behavior at the cellular and genetic levels; however, no reference genome exists for this species. To enable our studies, we are currently working toward assembling a reference-quality genome with chromosome level resolution. We are leveraging several sequencing technologies, including Illumina short read sequencing, Oxford Nanopore long read sequencing, PacBio long read sequencing, and both Chicago and Dovetail Hi-C sequencing to improve scaffolding. We will use this information to identify and target genes for genetic manipulation that are relevant to web behavior. We have sequenced and assembled a comprehensive transcriptome, and are currently working toward adapting RNAi protocols, developed in the spider *Parasteatoda tepidariorum* for maternal mRNA knockdown in embryos, for the purpose of knocking down behaviorally-relevant genes in behaving adults. We are also interested in characterizing other aspects of prey capture behavior by Uloborid spiders, including venomomics and silkomics, as well as addressing the evolution of unique aspects of prey capture behavior in Uloborids relative to other orb-weavers.

PO0485: Other Species

PhycoCosm: New JGI Resources for Algal Genomics

Richard D Hayes, Alan Kuo, Sara E. Calhoun, Igor Shabalov and Igor V. Grigoriev, US Department of Energy Joint Genome Institute, Berkeley, CA

Algae are a diverse, polyphyletic group of eukaryotes that are collectively responsible for over 50% of photosynthesis on Earth. However, the algal genetic blueprint is known for only a very small number of representative species, limiting our understanding of their complex biology and slowing advances in bioenergy, bioproduct development, and biomanufacturing. Therefore, the DOE Joint Genome Institute (JGI) has launched an algal genomics program, to promote research into the evolution, physiology, and molecular biology of these Eukaryotes, and to support DOE's Biological and Environmental Research program mission focus on bioenergy, the global carbon cycle, and biogeochemistry. In particular, algae are one of the focus areas of the latest Community Science Program call for proposals for high-throughput sequencing and other -omics projects (<https://jgi.doe.gov/user-programs/program-info/how-to-propose-a-csp-project/>).

In the last few years, the JGI has sequenced, annotated, and published a number of high-profile algal genomes, several of which were the first sequenced representatives of their respective phyla. PhycoCosm (<https://phycocosm.jgi.doe.gov>) has collected these JGI projects, together with a thorough survey of publicly available nuclear genome annotations, to produce a comprehensive algal genomics resource. Genome annotations are organized into individual portals, each with web-based analysis tools including a genome browser, synteny views, gene families built with comparative species, and targeted functional annotations. Our aim is to be an indispensable hub for algal comparative genomics, encouraging community engagement and data exchange, and fostering new sequencing projects that will further these research goals.

PE0486: Other Species

Biosynthesis of Poly-(3)-Hydroxyalkanoic Acid by *Bacillus megaterium* SF4 Isolated from Sugarcane Farm in South-Western, Nigeria

Temitope O. Fadipe^{1,2}, Nazia Jamil² and Adekunle K. Lawal³, (1)Federal Institute of Industrial Research, Oshodi, Lagos, Lagos, Nigeria, (2)University of the Punjab, Lahore, Pakistan, (3)Federal Institute of Industrial Research, Oshodi, Lagos, Nigeria

Plastics have formed an integral part of human lives with about 25 million tons produced annually. However, the non-degradability and subsequent persistence of synthetic polymers used in plastic production in the environment have been associated with ecosystem imbalance and health problems. This necessitates the development of eco-friendly and biodegradable polymers as alternatives in plastic production. Poly-(3)-hydroxyalkanoic acid (PHAs) are microbially derived biopolymers produced by bacteria under unbalanced growth conditions. The physical and chemical properties of PHAs are comparable with synthetic polymers used in petroleum-based plastics thus making them suitable alternatives in plastic production. We share results from the use of a bacterial isolate with potential for sustainable PHAs production. The isolate was screened for PHAs production in a nitrogen limiting medium (NLM) with 2% carbon source (glucose / glycerol / starch / sugarcane molasses) supplemented 0.5 µg/ml of Nile Red and Nile Blue A for fluorescence detection of PHAs and 0.3% Sudan Black B for intracellular detection of PHAs. The bacterium was biochemically characterized and further identified by 16SrRNA sequencing. Poly-(3)-hydroxyalkanoic acid production was carried out using pure culture of the isolate in a NLM with 2% carbon substrate (glucose / glycerol / starch / sugarcane molasses) over a 96 hour incubation period. Extraction of PHAs from lyophilized bacterial biomass was done by sodium hypochlorite/chloroform method and the extracted PHAs was subsequently characterized by FT-IR and GC-MS. The PHA synthase genes, *PhaC* and *PhaR* of the isolate were also partially amplified and sequenced. The isolate was identified as *Bacillus megaterium* SF4 and produced orange and yellow fluorescence for all carbon substrates used thus indicating the presence of PHAs. Blue black intracellular inclusions of PHAs were also observed after Sudan Black B staining. Growth curves showed highest biomass accumulation in 2% starch followed by 2% glucose and 2% glycerol respectively with least biomass accumulated in 2% sugarcane molasses. *Bacillus megaterium* SF4 produced PHAs of up to $26.53 \pm 0.91\%$ in starch, $4.69 \pm 0.17\%$ in glucose, $18.64 \pm 0.65\%$ in glycerol and $10.22 \pm 0.32\%$ in sugarcane molasses. FT-IR spectra showed peaks corresponding to the presence of P3HB and P3HB3HV. Detection of peaks at 1721 cm^{-1} reveal conformational changes in crystalline and amorphous phases in the extracted PHAs. Homology search of the 16SrRNA, *PhaC* and *PhaR* sequences revealed maximum identity with *B. megaterium* NBRC 15308. Sequences of 16SrRNA, *PhaC* and *PhaR* have been deposited in the NCBI GenBank repository with accession numbers KY855376.1, KY855378.1 and MF947449.1 respectively. The results show that *B. megaterium* SF4 has great potential for sustainable PHAs production.

PO0487: Other Species

Differences in the Genomes of *Phyllosticta Citricarpa* and *Phyllosticta Capitalensis*.

Marco Aurélio Takita¹, Carolina Munari Rodrigues¹, Nicholas V. Silva², Marcelo Ribeiro-Alves³ and Marcos A. Machado¹, (1)Citrus Center, IAC, Sao Paulo State Secretary of Agriculture and Supply, Cordeiropolis, SP, Brazil, (2)Universidade Estadual de Campinas, Campinas, Brazil, (3)Instituto Nacional de Infectologia Evandro Chagas, Fundação Oswaldo Cruz, Rio de Janeiro, Brazil

Citrus are affected by many diseases that are caused by different pathogens. Besides, Citrus also hosts many symbiotic microorganisms in a relationship that may be advantageous for both organisms. The fungi *Phyllosticta citricarpa* is a pathogen that is responsible for citrus black spot, and with *Phyllosticta capitalensis*, an endophytic species, are examples of closely related species with different behaviour in citrus. Both species are biologically associated and share a very similar morphology, and in order to identify genetic differences that could explain their lifestyles, genomes sequencing were carried.

Drafts genomes were assembled with sizes close to 33 Mb for both fungi. They carry 15,206 and 14,797 coding sequences for *P. citricarpa* and *P. capitalensis*, respectively. Enrichment analysis shows that the pathogenic species presents growth and development genes that may be necessary for its pathogenicity. On the other hand, family expansion analyses showed the plasticity of the genome of these species.

Genome evolution seems to be of real importance among the *Phyllosticta* isolates and it is leading to different biological characteristics of these species.

PE0488: Other Species

A New, Chromosome-Level *Aspergillus Flavus* Reference Genome Reveals Large Insertions Potentially Contributing to Isolate Stress Tolerance and Aflatoxin Production

Jake Fountain¹, Josh Clevenger², Brian Nadon³, Justin N Vaughn⁴, Ramey C. Youngblood⁵, Walid Korani⁶, Perng-Kuang Chang⁷, Gaurav Agarwal¹, Bob Kemerait⁸, Manish K Pandey⁹, Deepak Bhatnagar¹⁰, Peggy Ozias-Akins¹, Rajeev K Varshney¹¹, Brian E. Scheffler¹² and **Baozhu Guo**¹, (1)University of Georgia, Tifton, GA, (2)Mars Wrigley Confectionery, Athens, GA, (3)Genomics and Bioinformatics Research Unit, USDA-ARS, Athens, GA, (4)USDA-ARS/University of Georgia, Athens, GA, (5)Mississippi State University, Mississippi State, MS, (6)University of Georgia, Athens, GA, (7)SRRC/ARS/USDA, New Orleans, LA, (8)University of Georgia, TIFTON, GA, (9)ICRISAT, Hyderabad, India, (10)SRRC, (11)ICRISAT, Patancheru, India, (12)USDA-ARS, Stoneville, MS

Previous efforts made in genome sequencing in the *Aspergillus* genus have led to the development of high quality reference genomes for several important species including *A. nidulans*, *A. fumigatus*, and *A. oryzae*. However, less progress had been made with regard to the agriculturally important species *A. flavus*. Efforts began in 2001 to sequence the genome of the isolate NRRL3357 resulting in a scaffold-level genome released in 2005. Since then, the isolate AF70 was sequenced to the scaffold level using an Illumina approach in 2015, and the NRRL3357 isolate was re-sequenced and assembled to chromosome lengths using long-read approaches in 2019. While these resources provide a base for genomics research in *A. flavus*, additional assemblies coupled with comparative and phylogenetic analyses will provide new insights into differences among isolates related to mycotoxin production, pathogenicity, and other phenotypic traits. To explore these important traits and the underlying genetic variation contributing to their diversity, here we present a new, chromosome-level reference genome for AF13, a MAT1-2, highly stress tolerant, and plant pathogenic isolate of *A. flavus*, and a comparative analysis with a complete, chromosome-level assembly of NRRL3357. This analysis resulted in the identification of 153 and 45 unique genes in AF13 and NRRL3357, respectively. Structural variation analyses coupled with optical mapping also confirmed the presence of a large 310Kb insertion present in AF13 containing 58 genes unique to the isolate. Analysis of this insertion revealed the presence of a bZIP transcription factor, *atfC*, which may contribute to isolate pathogenicity and stress tolerance. Additional pathogenicity and stress tolerance-related unique genes were also observed within the insert and elsewhere in the AF13 genome, and may provide an explanation for the greater level of stress tolerance observed in AF13 compared to NRRL3357 and other isolates of *A. flavus*.

PO0489: Other Plant Species

Development and Validation of High-Throughput PACE Markers in Cannabis sativa L.

Jacob Toth¹, George M. Stack¹, Ali R. Cala², Craig H. Carlson¹, Rebecca L. Wilk¹, Jamie L. Crawford³, Donald R. Viands³, Glenn Philippe⁴, Christine D. Smart², Jocelyn K.C. Rose⁴ and Lawrence B. Smart¹, (1)Horticulture Section, School of Integrative Plant Science, Cornell University, Geneva, NY, (2)Plant Pathology and Plant-Microbe Biology Section, School of Integrative Plant Science, Cornell University, Geneva, NY, (3)Plant Breeding and Genetics Section, School of Integrative Plant Science, Cornell University, Ithaca, NY, (4)Plant Biology Section, School of Integrative Plant Science, Cornell University, Ithaca, NY

Hemp (*Cannabis sativa* L.) is an emerging dioecious crop grown primarily for grain, fiber, and cannabinoids. There is good evidence for medicinal benefits of the most abundant cannabinoid in hemp, cannabidiol (CBD). For CBD production, female plants producing CBD but not tetrahydrocannabinol (THC) are desired. We developed and validated high-throughput PACE (PCR Allele Competitive Extension) assays for *C. sativa* plant sex and cannabinoid chemotype. The sex assay was validated across a wide range of germplasm and resolved male plants from female and monoecious plants. The cannabinoid chemotype assay revealed segregation in hemp populations, and resolved plants producing predominantly THC, predominantly CBD, and roughly equal amounts of THC and CBD. Cultivar populations that were thought to be stabilized for CBD production were found to be segregating phenotypically and genotypically. Many plants predominantly producing CBD accumulated more than the current legal limit of 0.3% THC by dry weight. These assays and data provide potentially useful tools for breeding and early selection of hemp.

PE0490: Other Plant Species

Development of a High-Throughput SNP Panel for Marker-Assisted Precision Breeding in Cannabis

Kathleen McGrath, Dr. Elie Raherison, Dr. Farzad Javidfar, Dr. Shivappa Hukkeri, Dr. Dinesh Adhikary, Max Hargreaves, Michael Lowings and Dr. Manoj Kulkarni, Destiny Bioscience, Nisku, AB, Canada

As Cannabis breeding expands into the medicinal sector, there is an urgent need for increased genetic and genomic resources that will capture the high level of genetic diversity and heterozygosity in Cannabis. The ability to link with certainty genotypes to phenotypes and chemotypes is integral to precision Cannabis breeding. Additionally, the complexity of polygenic inheritance for cannabinoid, terpene, and flavonoid biosynthesis genes makes it difficult to select while breeding using only chemotype data in the absence of molecular markers.

High-throughput SNP genotyping provides a scalable and cost-efficient method to identify known and novel chemotypes with the ability to match putatively beneficial germplasm in a breeding pool. Present efforts in Cannabis marker-assisted breeding include some SNPs for cannabinoid, terpene, and flavonoid biosynthesis genes but lack comprehensive genome coverage and other yield improvement traits (agronomical, physiological, and disease resistance). Destiny is utilizing in-house and publicly available databases to establish a high-throughput targeted precision breeding panel with 90,000 SNPs covering the entire genome. This breeding panel will be further validated using Illumina's Infinium iSelect array and iScan platform, and available genetic diversity for substantiating its use in Cannabis breeding. This marker-assisted, chemotype based breeding is an integral step towards personalized medical Cannabis formulations, allowing for the type of precision and confirmation rigor that is currently lacking in Cannabis breeding research.

P00491: Other Plant Species

Genome Assembly of Apomictic, Polyploid Kentucky Bluegrass (*Poa pratensis* L.)

Bradley S. Bushman¹, Matthew D. Robbins², Irina Orlova³ and Bob Harriman³, (1)USDA-ARS Forage & Range Research, Logan, UT, (2)USDA ARS Forage and Range Research Lab, Logan, UT, (3)Scotts Miracle Gro

Kentucky bluegrass a common cool-season (C3) turfgrass for lawns, parks, and other uses in temperate climates; and its wide adaptability has contributed to a worldwide distribution in unmanaged areas. The species occurs as a high polyploid complex, between 8-14x, with abundant mitotic and meiotic aneuploidy. The progenitor species are unknown and may no longer exist, and the species is hypothesized to be partially auto-polyploid. With the self-incompatibility system of Poaceae grasses, the heterozygosity of Kentucky bluegrass is relatively high. Furthermore, as a facultative apomict, sexual offtypes occur at varying frequencies such as canonical hybrids, polyhaploids, and fertilization events without meiosis. In this context, we selected a ~3.5Gb polyhaploid of the cultivated variety 'Hampton' and sequenced on a PacBio Sequel using the CLR approach. The raw read N50s ranged from 19kb to 37kb, with 470Gb of total sequence generated. Canu v1.8 was applied for *de novo* genome assembly, with parameters optimized for accuracy and polyploidy with a corrected error rate of 0.085. The resulting assembly was a total of 4.2Gb in 4365 contigs with an NG50 of 22.3Mb. Contigs from the assembly were twice polished with Arrow (SMRT Link v7.0.0, resequencing2 pipeline v0.2.0), and once with Pilon v1.23 using paired end Illumina reads. Busco v3, using the Liliopsida dataset, had 96.4% complete coverage, with the vast majority of genes having 2-3 hits.

PE0492: Other Plant Species

Whole Genome Sequencing of 43 Cannabis Cultivars Identifies Putative Powdery Mildew Resistance Gene.

Kevin McKernan¹, Yvonne Helbert¹, Liam T Kane², Heather Ebling², Lei Zhang², Biao Liu², Zachary Eaton², Luo Sun³, Brittany S Sexton³, Keerthana Krishnan³, Eileen Dimalanta³, Sarah Kingan⁴, Primo Baybayan⁴, Gregory T Concepcion⁴, Brad Barbazuk⁵ and Timothy Harkins², (1)Medicinal Genomics, Beverly, MA, (2)Medicinal Genomics, Woburn, MA, (3)New England Biolabs, Ipswich, MA, (4)Pacific Biosciences, Menlo Park, CA, (5)University of Florida, Gainesville, FL

We whole genome sequenced and assembled a cannabis trio utilizing the Pacific Biosciences Sequel II platform. This trio was complemented with 40 Whole genome sequences (50X) from various hemp and cannabis cultivars using the Illumina Novaseq platform. Copy Number Variation analysis was utilized to identify amplified and deleted

genes in the cannabis genome related to pathogen response. RNA expression information was collected utilizing Pacific Bioscience's IsoSeq method. Five male and female mRNA tissue types were analyzed to annotate the genome. 27,664 and 32,108 genes were identified in Female and Male genomes respectively. These annotations were utilized in conjunction with copy number analysis to prioritize 23 putative pathogen response genes in structural variation amongst the cultivars. Quantitative RT-PCR assays were developed targeting *Golovinomyces chitoracearum* and pathogen response genes in Powdery Mildew infected and Powdery Mildew resistant cannabis cultivars. The strongest candidate gene was cloned into pET vectors for bacterial expression and fungicidal *in-vitro* assays. Powdery mildew is a destructive cannabis and hemp pathogen that can result in reduced crop yield. OSHA and the CDC have reported it as a human allergen in cannabis trimming environments. Identification of the resistance markers can accelerate resistance breeding programs.

PO0493: Other Plant Species

Evaluation of 30 CBD Hemp Cultivars in New York State

George M. Stack¹, Jacob Toth¹, Craig H. Carlson¹, Mariana Marrero-González¹, Rebecca Wilk¹, Deanna Gentner¹, Jamie Crawford², Glenn Philippe³, Jocelyn KC Rose³, Donald Viands² and Lawrence B. Smart¹, (1)Horticulture Section, School of Integrative Plant Science, Cornell University, Geneva, NY, (2)Plant Breeding and Genetics Section, School of Integrative Plant Science, Cornell University, Ithaca, NY, (3)Plant Biology Section, School of Integrative Plant Science, Cornell University, Ithaca, NY

Hemp (*Cannabis sativa* L. <0.3% THC) is an emerging crop in New York State due to its multitude of uses and suitability to grow in the local climate. One barrier to rapid adoption is the lack of available yield and quality data about commercial cultivars as a result of the several-decade lapse in production. We characterized 30 CBD-type hemp cultivars for both immediate use by growers and potential value in a breeding program. To quantify the phenotypic diversity of these cultivars, we collected data throughout the lifetime of the plants – from seed germination to post-harvest processing – across two sites in New York. Many available seed-propagated cultivars still segregate for important traits like form, cannabinoid profile, and flowering time. All cultivars showed phenotypic variation by site. We found significant variation in yield-determining traits including growth rates, whole plant biomass, floral tissue biomass, flowering time, and cannabinoid profile.

PE0494: Other Plant Species

Genetic Diversity of Cannabidiol (CBD) Rich Industrial Hemp Accessions

Matthew Johnson, The University of Georgia, Athens, GA

Industrial hemp acreage in the US has been rapidly growing every year since 2014. This rapid growth has left many farmers rushing to plant as many acres as they can to cash in on this opportunity. Unfortunately, many of these farmers have lost millions of dollars by purchasing low quality seed. These seeds are a result of the lack of professional breeding since industrial hemp was legalized in the US. Even seeds that are not causing farmers to lose their crops do not have the levels of consistency farmers have grown accustomed to, and plants within a field have different harvesting times, heights, general architecture, and cannabinoid content. These inconsistencies can all cause issues for farmers, but the inconsistency in cannabinoid contents are the greatest risks, since plants which get above 0.3% THC are no longer in compliance with federal regulations and might cause the fields to be destroyed. To better understand the issue of seed inconsistencies, we are using genotyping-by-sequencing to generate markers in diverse germplasm of high cannabidiol (CBD) industrial hemp accessions and analyze the genetic diversity within and between each accession. Seeing the extent of genetic diversity within a given accession will give us a better understanding of how it might affect farmers who are planting these seeds. This information will also serve as a useful tool for breeding industrial hemp strains, and will enable hemp breeders to take full advantage of genetic diversity to breed improved crops.

PO0495: Other Plant Species

Phenotypic and Genetic Analysis of Hemp (*Cannabis sativa* L.) Cultivars Grown in New York State

Craig H. Carlson¹, Jamie Crawford², George M. Stack¹, Ali Cala³, Glenn Philippe⁴, Jacob Toth¹, Jocelyn KC Rose⁴, Christine D. Smart³ and Lawrence B. Smart¹, (1)Horticulture Section, School of Integrative

Plant Science, Cornell University, Geneva, NY, (2)Plant Breeding and Genetics Section, School of Integrative Plant Science, Cornell University, Ithaca, NY, (3)Plant Pathology and Plant-Microbe Biology Section, School of Integrative Plant Science, Cornell University, Geneva, NY, (4)Plant Biology Section, School of Integrative Plant Science, Cornell University, Ithaca, NY

Hemp (*Cannabis sativa* L., < 0.3% total potential THC) is a multi-purpose annual crop consisting of distinct ideotypes that specialize in grain, fiber, and CBD (cannabidiol) production. The market potential for hemp and its products, including food, personal care, health and supplements, animal feed, textiles and industrial products is estimated to grow to a multi-billion dollar industry within the decade. Commercially available cultivars from Europe and Canada have been bred for enhanced production of oil, fiber, and CBD, but have limited or no testing in NY. Because many available hemp cultivars were developed in environments very different than NY, it is critical to test hemp cultivars to determine which are well adapted and will be high yielding across and within NY State's range of soil types and environments. Phenotypic data (e.g., yield, cannabinoid profiles, flowering time, and disease resistance) from grain, fiber, and CBD cultivar trials was collected across multiple environments and years. While the results of these trials are primarily reported to better inform hemp growers and processors, they are also particularly useful to the breeder, like identifying parents to be used in crosses. In addition to assessing phenotypic variance, it is critical that genetic relatedness or kinship among cultivar populations is known, so breeding decisions can be better informed. From a total of 192 grain, fiber, dual-purpose, and CBD cultivars, as well as feral US accessions, DNA was isolated and genotyped using a genotyping-by-sequencing (GBS) approach. Utilizing genome-wide variants obtained from GBS alignments to the latest hemp genome reference, heterozygosity, kinship, and population structure was determined using various statistical techniques. In addition, a subset of these individuals were used in phenotype ~ genotype associations for cannabinoid content and seed-related traits. Bi-parental genetic mapping populations are already being developed using parents selected from these cultivar populations for important traits related to yield and composition. With continued phenotypic and genetic assessment of hemp, there is vast potential to make dramatic gains in this exciting new crop.

PE0496: Other Plant Species

***Cannabis sativa* Genotyping and Exon Sequencing using a New Jamaican Lion-Derived Hybridization Capture Panel**

Jacob Enk¹, Alison Devault¹, Biao Liu² and Kevin McKernan², (1)Arbor Biosciences, Ann Arbor, MI, (2)Medicinal Genomics, Beverly, MA

Second and third-generation sequencing technologies have recently been leveraged to resolve genomes of a variety of cannabis cultivars. Using this tremendous resource, Medicinal Genomics Corporation (MGC) and Arbor Biosciences have partnered to develop a new hybridization capture panel for comprehensive, routine sequencing of key cannabis genomic loci. Comprising 8.4 Mbp total target space, the kit consistently retrieves not only tens of thousands of SNP markers for high resolution genotyping, but also regions involved in cannabinoid and terpene synthesis, as well as edestin and seed production. The co-branded panel will be available as a kit from MGC for the global cannabis breeding market, and from Arbor Bio for the global hemp research market. This cost-effective sequencing tool enables researchers to better understand genetic pathways responsible for favored traits, and growers, dispensaries and testing laboratories to ensure patients and consumers have access to safe, quality cannabis.

PO0497: Other Plant Species

Genomics of Oil Profile Variation in Cannabis

Keith Allen, 1959, Port Ludlow, WA

The availability of multiple good quality Cannabis reference genomes has made it possible to map the gene families for the enzymes that make cannabinoids, terpenes, and their precursors. We have mapped 55 terpene synthase genes, about 20 cannabinoid synthase genes, and a number of upstream pathway genes in several genome assemblies. The Cannabis terpene synthase family is extremely diverse, and is expressed in all parts of the plant, with most genes having substantial tissue specificity. The cannabinoid synthase family is also expressed throughout the plant, including the roots. In both of these families some of the genes are in clusters of varying degrees of compactness. Interestingly, we find variation in the structure of some of these clusters between genome assemblies from different cultivars, suggesting that active rearrangement is still going on. Many of these genes are not functionally

characterized, and it is the nature of these enzymes that exact function (ie, product profile) cannot be inferred from primary sequence, so we have a ways to go in fully understanding the genetic basis for oil profile variation in Cannabis. A full description of the genomic arrangement, expression variation and sequence variation is an important step in getting us to that full understanding.

PE0498: Other Plant Species

Leveraging Genome-Wide Association and Linkage Mapping Approaches to Identify Loci for Resistance to Hop Powdery Mildew in *Humulus lupulus* L.

Joshua S. Havill¹, Casey J. McNichols¹, Adam L. Schrankler¹, John Henning² and Gary J. Muehlbauer³, (1)University of Minnesota, St. Paul, MN, (2)USDA-ARS, Corvallis, OR, (3)Department of Agronomy and Plant Genetics, University of Minnesota, St. Paul, MN

Hop powdery mildew, caused by *Podosphaera macularis*, is a barrier to common hop (*Humulus lupulus* L.) production in the Northern hemisphere. Recently, outbreaks of virulent pathogen strains overcoming widely-deployed resistant varieties has prompted the search for novel resistance genes or combining multiple extant resistance genes to combat this disease. Use of molecular markers has the capacity to increase the efficiency of breeding in this perennial cropping system and potentially lead to the identification of genes responsible for conferring resistance. Herein, we describe the development of three populations of three-hundred and eighty-four individuals (one-thousand one-hundred and fifty-two seedlings in total) each representing various hop subspecies (*H. l.* var. *lupuloides*, *H. l.* var. *lupulus*, and *H. l.* var. *neomexicanus*) from North America or Eurasia, which were originally selected from 6,149 seedlings exhibiting varying degrees of resistance among subspecies. Resistant seedlings were identified in the following frequencies among the subspecies 2.57% in *H. l.* var. *lupuloides*, 9.26% in *H. l.* var. *lupulus*, and 4.39% in *H. l.* var. *neomexicanus* prior to down-sampling within each population. Genotyping these populations is in process and these data will be combined with the phenotype data to conduct genome wide association mapping. Additionally, mapping powdery mildew resistance in a bi-parental population from a cross between the resistant hop cultivar 'Zenith' and the susceptible male breeding line '21058M' will be presented.

PO0499: Other Plant Species

Insights into Genetic Structure and Genomic Prediction of Powdery Mildew Resistance in Flax (*Linum usitatissimum* L.)

Sylvie Cloutier¹, Frank You¹, Khalid Rashid², Chunfang Zheng³, Samuel Lam³, Nadeem Khan³, Liqiang He³, Jin Xiao⁴, Zhen Yao² and Pingchuan Li², (1)Ottawa Research and Development Centre, Agriculture and Agri-Food Canada, Ottawa, ON, Canada, (2)Agriculture and Agri-Food Canada, Morden, MB, Canada, (3)Agriculture and Agri-Food Canada, Ottawa, ON, Canada, (4)Key Laboratory of Crop Genetics and Germplasm Enhancement, College of Agriculture, Nanjing Agricultural University, Ottawa, ON, Canada

In flax, powdery mildew (PM) is caused by the fungus *Oidium lini* which can cause defoliation and reduce seed yield and quality. Until now, only one major dominant gene (*Pm1*) and three QTL on chromosome 1, 7 and 9 were reported for PM resistance. To fully dissect genetic structure and identify quantitative trait loci for PM resistance, a diverse flax core collection of 372 accessions and an additional 75 breeding lines were re-sequenced and field resistance evaluation was performed for six to eight years (2010-2017) in Morden, Manitoba, Canada. Genome-wide association studies were performed using two single-locus and seven multi-locus statistical models, 247,160 single nucleotide polymorphisms (SNPs) and the phenotypes of the 447 individuals for each individual years and the means over years. A total of 355 QTL based on haplotype blocks were identified, of which, 43 large effect QTL ($\geq 10\%$ of R^2) were highly stable over years. All QTL explained 91% of the phenotypic variation in the genetic panel. The total number of favorable alleles per accession was significantly correlated with PM resistance ($r = 0.74$) and genomic selection (GS) models using all identified QTL generated significantly higher prediction accuracy ($r = 0.93$) than those using all genome-wide SNPs ($r = 0.42$), validating the overall reliability of the QTL. The QTL were found on all 15 chromosomes but the 43 major QTL were clustered in 16 genomic regions of 12 chromosomes, especially on chromosome 5 (0.4-5.6 Mb and 9.4-16.9 Mb) and 13 (4.7-5.2 Mb). Of the 355 QTL, 236 harbored disease resistance related genes (390) within 200 Kb. In particular, 27 major QTL co-located with NBS-related genes, receptor-like kinase, protein-like kinase, TM-CC, RPW8, WRKY, and mildew locus O (MLO) genes. In

conclusion, QTL for PM resistance are additive and GS models based on QTL data are effective for predicting resistance of breeding lines for use in flax molecular breeding.

PE0500: Other Plant Species

Tracing Histories of Polyploidy, Admixture, and Diversification in a Hawaiian Plant Radiation through Genomic Analyses

Crystal Tomlin¹, Sitaram Rajaraman², Jarkko Salojärvi², Todd Michael³, Victor A. Albert¹ and Charlotte Lindqvist¹, (1)University at Buffalo, Buffalo, NY, (2)Nanyang Technological University, Singapore, Singapore, (3)J. Craig Venter Institute, La Jolla, CA

Island systems are among the most important natural environments for studying fundamental processes underlying species diversification, migration, and extinction. The Hawaiian endemic mints (Lamiaceae) display a remarkable diversity, comprising 59 species in three genera (*Phyllostegia*, *Stenogyne*, and *Haplostachys*), making this group an ideal model to study adaptive radiation and speciation from a genomic scale. Previous research has suggested a complex hybrid and octoploid origin for this lineage, but plastid genomes and limited nuclear data have been unable to determine a detailed evolutionary history in the face of rampant gene flow and rapid evolutionary divergence. In this study, we are assembling the genome of a Hawaiian mint species, *Stenogyne calaminthoides*, using Oxford Nanopore and Illumina sequencing technology, and HiC scaffolding. However, a relatively large genome size of ~1.4 Gb, a substantial proportion of very recently proliferated LTR retrotransposons, and high heterozygosity make *de novo* assembly challenging. Additionally, we resequenced 46 Hawaiian mints, including their New and Old World close relatives, and mapped sequences to a draft reference genome of *S. calaminthoides*. Through phylogenetic analyses, we find support for monophyly of each Hawaiian mint genus and a close relationship between Hawaiian mints and North American west coast *Stachys* species. However, tests for admixture indicate extensive gene flow among Hawaiian mints and we find an increase in effective population size in recent times for admixed samples. Genome level analysis is providing a lens into the evolutionary past of the Hawaiian mint radiation, suggesting an even deeper level of complexity than initially expected.

PO0501: Other Plant Species

The Tuareg Pharmacopeia: Basil, Aloe and Willow Teas Work on Drug Resistant Cancers

David Lightfoot, Southern Illinois University, Carbondale, IL

The original aim of the analysis was based on willow leaves. Aspirin was less effective than willow tea leaves? Basil tea worked the best. Ethanol extracts were less effective. Surprisingly, of just 13 natural products of plants used by the Tuaregs we have predicted the most likely artificial mixtures of 2-3 most effective natural products on drug resistant leukemia cells from over 364 possible mixtures. The natural products selected included resveratrol, honokiol, chrysin, limonene, cholecalciferol, cerulenin, aloe emodin, and salicin. They had over 600 potential protein targets. Target profiling used the OntomineTM set of tools for literature searches of potential binding proteins, binding constant predictions, binding site predictions, and pathway network pattern analysis. The analyses indicated that 6 of the 13 natural products predicted binding proteins which were important targets for established cancer treatments. Improvements in effectiveness were predicted for artificial combinations of 2 or 3 natural products. That effect might be attributed to drug synergism rather than increased numbers of binding proteins bound (dose effects). Among natural products, the combinations of Aloe emodin with mevinolin and honokiol were predicted to be the most effective combination for AML-related predicted binding proteins. Therefore, plant extracts may in future provide more effective medicines than the single purified natural products of modern medicine, in some cases. However, drink herbal teas whenever you can...

PE0502: Other Plant Species

Antagonistic Regulation of Auxiliary Bud Outgrowth By the Branched Genes in Tobacco

Shengming Yang, USDA, Fargo, ND

As a key signaling integrator of shoot branching, *BRANCHED 1 (BR1)* coordinates and is orchestrated by endogenous and environmental signals involved in the regulation of axillary bud outgrowth. Using CRISPR site-

directed mutagenesis and overexpression assays, we characterized the regulatory roles of five BRC gene members in tobacco. We show that lateral branching is negatively regulated by *NtBRC1A-1*, *1B-1*, and *1B-2*, but was unexpectedly promoted by *NtBRC2A*. The *NtBRC1A-2* gene seems not to be required for the regulation of axillary bud outgrowth. Suppression of bud growth may be attained by direct binding of NtBRC proteins to *Tassels Replace Upper Ears 1 (TRUI)* genes, as evidenced by RNAi and CRISPR-mediated gene editing. The *BRC2* gene, which is thought to be unnecessary for branch development, probably confers a dominant negative effect by interfering with the branching-inhibitory *BRC1* genes. However, the molecular mechanism of NtBRC2A action still needs to be further addressed. Our results may indicate that there is a blind spot when using genetic tools, such as CRISPR and RNAi, in the functional characterization of members of a gene family.

PO0503: Other Plant Species

Automated High-Throughput Quantitative Phenotyping of Boxwood Blight

Fred E. Gouker, USDA-ARS, Beltsville, MD

Quantitative phenotyping is frequently used in plant breeding and genetic studies, as well as in studies focused on pathogen biology such as chemical efficacy trials. In these scenarios, phenotyping a large number of genotypes or treatments can be advantageous but is often limited by time and cost. Here we present a novel computational pipeline dedicated to estimating the percent area of boxwood blight infection and sporulation from images of inoculated boxwood leaf discs that is accurate and time efficient. The pipeline was tested on images from leaf disc assay experiments involving 47 interspecific F1 boxwood families. Correlations between computer vision and manual visual ratings reached 0.87 across all families. Additionally, this data was used to train and develop a machine learning protocol to estimate percent infection and sporulation that was better than either manual or computer vision ratings. It is estimated that implementation of this machine learning phenotyping method would use at least 80% less time than using the computer vision system or the manual rating method. This will allow more treatments to be phenotyped in order to better understand the genetic architecture of boxwood blight resistance.

PE0504: Other Plant Species

Characterization of Seven Polymorphic Genes Controlling Red Leaf Color in Lettuce and Their Applications in Lettuce Improvement

Rong Tao^{1,2}, Wenqing Su², Wenye Liu², Guanghui An², Lei Zhang², Weiyi Zhang², Richard Michelmore¹, Jiongjiong Chen² and Hanhui Kuang², (1)The Genome Center, University of California, Davis, CA, (2)Huazhong Agricultural University, Wuhan, China

Lettuce (*Lactuca sativa*) is one of the most important vegetable crops worldwide and shows dramatic variation in morphology and leaf color. Anthocyanin accumulation in lettuce was one of the first traits to be analyzed genetically in plants. Using genome-wide association studies, eQTLs, gene expression network as well as bi-parental segregation analysis, we identified seven loci controlling variation in leaf color of lettuce, and cloned five of them, which encode R2R3-MYB, R3-MYB, bHLH, WD40 and ANS, respectively. Mutations in the *bHLH* or *ANS* genes completely abolished anthocyanin biosynthesis, resulting in pure green leaves. In contrast, mutations in the other three genes promoted anthocyanin accumulation, revealing of disruptive selection for leaf color of lettuce. Another two loci were fine mapped to a small region, and one of them originated from a spontaneous mutation discovered in our field, and the causal mutation is currently under investigation. Pyramiding color-promoting alleles of the seven loci generates dark red phenotype, and different combinations of alleles of these seven genes may generate different levels of anthocyanin. Green cultivars with high levels of colorless flavonoids may be produced when mutated *ans* gene is selected. Green cultivars with rich flavonoids and red cultivars with different color intensity would provide health-promoting lettuce products to consumers with diverse preferences of the color of lettuce leaves.

PO0505: Other Plant Species

A New Approach to Control Vernalization Requirement in Biennial Plants

Yosuke Kuroda, NARO, Hokkaido Agricultural Research Center(HARC), Hokkaido, Japan

A biennial plant takes two years to flower and complete its lifecycle. In the first year, the plant grows vegetatively, then it enters a period of cold winter season. In the second year, it grows reproductively and initiates flower under long-day condition. This absolute vernalization requirement is a decisive factor in the lifecycle of a biennial plant. Interestingly, however, two strains have been found in biennial sugar beet (*Beta vulgaris*) that can flower quickly under long-day conditions without being exposed to cold temperatures. We named the hypothetical locus related to the flowering trend as 'B_{LOND}' (bolting by longer than natural daylength) and made crossing between strains of the B_{LOND} and normal biennial to obtain segregating generations of F₁, F₂ and BC₁F₁. Their flowering rates were investigated under a 24-hour day length by supplementing light at night both in summer and winter. Those results suggested that they were controlled by a few strong genes with dominant effect. Thus, the flowering tendency of B_{LOND} can be applied for rapid generation scheme.

This research was supported JSPS KAKENHI Grant Number JP 17K07616

PE0506: Other Plant Species

Abundance and Insertional Polymorphism of Carrot Mites and Demography of *Daucus Carota*

Alicja Macko-Podgorni, Katarzyna Stelmach, Kornelia Kwolek and **Dariusz Grzebelus**, University of Agriculture in Krakow, Krakow, Poland

Miniature inverted repeat transposable elements (MITEs) are small non-autonomous DNA transposons ubiquitous in plant genomes, mobilized by their autonomous relatives. *Stowaway* MITEs are derived from and mobilized by elements from the *mariner* superfamily. They constitute a significant fraction of the carrot genome, however there is lack of comprehensive analysis of variation caused by *Daucus carota* *Stowaway* MITEs (*DcStos*), their association with genes and putative impact on the genome evolution.

We report on 14 families of carrot *Stowaways*, *DcStos*, jointly occupying ca. 0.5% of the host genome. We systematically mined 31 genomes of wild and cultivated *D. carota*, which yielded 18.5 thousand copies showing a remarkable insertion site polymorphism. The genomic distribution of *DcStos* differed with respect to the origin of host populations corresponding with the four major groups of *D. carota* (wild European, wild Asian, eastern cultivated, western cultivated). We showed that *DcStos* were associated with genes and occurred most frequently in 5' and 3' UTRs. Individual families differed in their propensity to reside in particular segments of the genic region. Most importantly, *DcSto* copies in the 2kb up- and downstream regions were more frequently associated with genes encoding transcription factors, suggesting their possible functional impact. More than 1.5% of all *DcSto* insertion sites comprised different copies in exactly the same position in different host genomes, indicating the existence of insertional hotspots. The *DcSto7b* family was much more polymorphic than the remaining families in the cultivated carrot. We showed a line of evidence pointing at its activity in the course of carrot domestication and identified *Dcmar1* as an active carrot *mariner* element and a possible source of the transposition machinery for *DcSto7b*. *DcSto* intron length polymorphisms (DeS-ILPs) detected substantial genetic diversity and, showing considerable discrimination power, may be exploited as a tool for germplasm characterization and analysis of genome relationships. *DcSto* insertions mined from eastern cultivated carrots were usually much less frequent than those mined from the reference genome, possibly reflecting a bottleneck at the origin of the western carrot gene pool.

PO0507: Other Plant Species

Genome-Wide Association Mapping of Tuber Flesh Oxidation and Dry Matter Content in a Water Yam (*Dioscorea alata*) Panel.

Angelot Paterne AGRE¹, Pierre Gatarira², Ryo Matsumoto³, Patrick Olusanmi Adebola², Robert Asiedu⁴ and Asrat Asfaw Amele⁵, (1)International Institute of Tropical Agriculture (IITA/IBADAN), Ibadan, Oyo, Nigeria, (2)IITA Ibadan Nigeria, Ibadan, Nigeria, (3)Japan International Research Center for Agricultural Sciences, Tsukuba, Japan, (4)International Institute of Tropical Agriculture, Ibadan, Nigeria, (5)IITA, Abuja, Nigeria

Yam (*Dioscorea* spp.) is a versatile tuber crop grown in the tropics and sub-tropics as a preferred staple with many nutritional and medicinal significances. Among the cultivated yam species, water yam (*D. alata*) is most widely distributed and known for its good agronomic and tuber quality traits. A panel of 100 accessions was genotyped

using DArTSeq and phenotyped for tuber dry matter content and tuber flesh oxidation at three different locations to identify a region of genome involved in controlling the variation for tuber quality traits that would facilitate the development of new cultivars of water yam with improved food quality. A population structure analysis using 7442 SNP markers covering the yam genome identified four sub-genetic groups with 70% admixture within them. Genome-wide linkage disequilibrium (LD) analysis demonstrated that the average LD was about ~6 kb. A marker-trait association analysis in a linear mixed model that involved four different gene actions: additive, general, dominance alternative and dominance reference with the admixture group as a covariate identified hotspot regions significantly associated with tuber dry matter content in chromosomes 15 and 19 that cumulatively explained 22.40% of the total phenotypic variation for the trait. Likewise, nine regions of the genome spreader across six different chromosomes (1, 2,3,13,15, and 17) showed statistically significant associations with the variation in tuber flesh oxidation. Gene annotation for the regions with significant marker-trait associations revealed the presence of Peptidase C1A (IPR012599) and DEAD/DEAH (IPR011545) genes, which were previously reported as responsible for oxidation in many plants. An additional 28 candidate genes were identified in the peak SNP sites (or adjacent to these sites) for both tuber dry matter and tuber flesh oxidation with unknown functions. These results have elucidated the genetic architecture of dry matter content and tuber flesh oxidation in yam and revealed the suitability of GWAS in the identification of SNP variants associated with tuber food quality traits potentially applicable in yam breeding programs.

Keywords: *Dioscorea alata*, GWAS, Dry matter, Tuber flesh oxidation, DArT sequencing, gene annotation.

Submitter's E-mail Address: P.Agre@cgiar.org

PE0508: Other Plant Species

Plant Sex Prediction in White Guinea Yam (*Dioscorea rotundata*)

Patrick Olusanmi Adebola¹, Angelot Paterne AGRE², Chidinma Nwachkwu³, Bunmi Olasanmi³, Queen Obi⁴, Adeyinka Adewumi⁴, Idris Adejumbi⁴, Jude Obidiegwu⁵, Emmanuel Nwachukwu⁶, Asrat Asfaw Amele⁷ and David De Koeyer⁸, (1)IITA Ibadan Nigeria, Ibadan, Nigeria, (2)International Institute of Tropical Agriculture (IITA/IBADAN), Ibadan, Oyo, Nigeria, (3)Department of Agronomy, Ibadan, Nigeria, (4)International Institute of Tropical Agriculture, Ibadan, Nigeria, (5)National Root Crops Research Institute (NRCRI), Umudike, Nigeria, (6)National Root Crops Research Institute, Umudike, Nigeria, (7)IITA, Abuja, Nigeria, (8)Agriculture and Agri-Food Canada, Fredericton, NB, Canada

Plant sex in yam crop is complicated, with a monoecious and dioecious flowering pattern expressed by different genotypes. Plant sex determination with visual observation at the time of anthesis alone during the growth period can be very lengthy and time-consuming. Hence, early plant sex prediction at the seedling stage through molecular markers will be of great importance for designing an efficient hybridisation plan by enabling the selection of parental clones with defined sex information. The objectives of this study were to optimize and validate the appropriate leaf sampling method for high-quality DNA extraction and predict yam plant sex at the early growth or seedling stage in *D. rotundata* genotypes using molecular markers. Five leaf tissue sampling and preservation methods: liquid nitrogen, silica gel, 95% ethanol, dry ice, and oven-drying before DNA extraction were assessed for quality DNA extraction. Plant sex prediction was attempted in one hundred and ninety (190) genotypes using Sp16 SNP and Dr-Actin marker at the seedling stage. The marker predicted yam plant sex was validated with the visual flower sex phenotype score at the blooming stage. Liquid nitrogen, silica gel, dry ice, and oven drying gave the best quality DNA for leaf sample preservation before DNA extraction. Plant sex prediction at seedlings stage via the sp16 marker revealed more ZW genotypes (female/monoecious phenotypes) in the studied materials than ZZ genotypes (male phenotype) with prediction accuracy of 81.50%. These results have highlighted the potential applicability of the SP16 marker for plant sex prediction in the white yam breeding program.

Keywords: *D. rotundata*, DNA quality, leaf sampling, molecular marker, phenotyping, sex prediction

PO0509: Other Plant Species

Phenotypic and Genetic Variability in a Panel of White Guinea Yam (*Dioscorea rotundata*) Genotypes Based on Joint Analysis for Morphological and Molecular Diversity

Asrat Asfaw Amele¹, Kwabena Darkwa², Angelot Paterne AGRE³, Bunmi Olasanmi⁴, Iseki Kotharo⁵, Adrian Powell⁶, Guillaume Bauchet⁷, Lukas Mueller⁷, Susan R. Strickler⁶, David De Koeper⁸, Ryo Matsumoto⁹, Satoru Muranaka⁵, Robert Asiedu¹⁰ and Ryohei Terauchi¹¹, (1)IITA, Abuja, Nigeria, (2)IITA, Ibadan, Nigeria, (3)International Institute of Tropical Agriculture (IITA/IBADAN), Ibadan, Oyo, Nigeria, (4)University of Ibadan, Ibadan, Nigeria, (5)JIRCAS, Tokyo, Japan, (6)Boyce Thompson Institute, Ithaca, NY, (7)BTI, Ithaca, NY, (8)Agriculture and Agri-Food Canada, New Brunswick, ON, Canada, (9)Japan International Research Center for Agricultural Sciences, Tsukuba, Japan, (10)International Institute of Tropical Agriculture, Ibadan, Nigeria, (11)Iwate Biotechnology Research Center, Kitakami, Japan

A better understanding of the nature and extent of phenotypic and genetic diversity among germplasm in a breeding program is vital for making progress with genetic improvement. This study assessed variability in a panel of white guinea yam (*D. rotundata*) genotypes using joint analysis for phenotypic and genotypic diversity. The efficiency of different analysis methods to dissect the diversity in a yam germplasm pool was assessed using 136,426 SNP markers and 23 morphological attributes. The degree of resemblance between the original distance and result from cluster configuration varied among the different dissimilarity matrices and hierarchical clustering methods. The average (UPGMA) method showed high goodness-of-fit with Gower, and the Identity by the state (IBS) distances as it produced the most significant cophenetic correction values. The diversity estimated among the 173 white yam accessions was higher with molecular than phenotypic data. The grouping of genotypes into useful clusters showed a high inconsistency between the phenotypic and molecular data due to the non-overlapping information among the dissimilarity matrices. However, joint analysis for phenotypic and molecular data produced higher diversity indices that good-enough to capture existing genetic variability in the yam germplasm pool. The results from our study have provided valuable insights to inform breeding strategies and identify promising divergent parents for the development of improved white yam varieties with acceptable end-user qualities.

PE0510: Other Plant Species

Genomic Prediction of Heading Date in Intermediate Wheatgrass (*Thinopyrum intermedium*)

Raja Ragupathy¹, James R. Larsen², Michele Frick¹, Douglas J. Cattani³, Jordan Miller¹, James Elwood¹, Ariel Zezula¹, Rodrigo Ortega-Polo¹, Braulio J. Soto-Cerda⁴, Francois Eudes¹ and André Laroche¹, (1)AAFC - Lethbridge Research and Development Centre, Canada, (2)AAFC - Harrow Research and Development Centre, Canada, (3)Department of Plant Science, University of Manitoba, Canada, (4)Centro de Genómica Nutricional Agroacuícola (CGNA), Temuco, Chile

In the changing climate scenario, perennializing annual crops is one of the prime strategies of sustainable crop production. Intermediate Wheatgrass (IW) is used as a source of 'perenniality' in crosses with annual wheat, and also being *de novo* domesticated as a new perennial cereal. Hence, there is a need to understand the genetic architecture of phenotypic traits contributing to perenniality. In this study, a diverse panel consisting of 276 individuals collected from The Land Institute and USDA-National Germplasm Resources Laboratory were evaluated for heading date at two Canadian locations, namely Lethbridge (2017, 2018 and 2019) and Winnipeg (2018). Phenotypic data was analysed using mixed model, and best linear unbiased predictor (BLUP) values of lines combined over years were estimated. Genotyping of accessions was carried by GBS and the sequencing data was processed by fast-GBS pipeline (Torkamaneh et al. 2017) using *Thinopyrum intermedium* draft assembly (Kantarski et al. 2017). Genotype calling and downstream quality filtering resulted in the identification of 73,401 genome-wide SNPs among 276 accessions. Population structure was studied using 18,763 random genome-wide SNPs in STRUCTURE v.2.3.4 (Pritchard et al. 2000) suggesting two populations with 243 and 33 accessions, respectively. The average coefficient of membership of individuals belonging to Population-1 was 0.872 with a percentage of shared alleles with Population-2 individuals of 0.128. The coefficient of population differentiation *F_{st}* was 0.12, which indicated a moderate population structure between populations. Five different genomic prediction models, namely ridge regression-best linear unbiased predictor (rrBLUP, Endelman 2011), Bayesian ridge regression (Bayesian rr; de los Campos et al. 2013), Bayes A and Bayes B (Meuwissen et al. 2001), and genomic best linear unbiased predictor (GBLUP; Habier et al. 2007) were evaluated with 80% of the lines as the training set, for estimating marker effects, and prediction accuracies ranged from 0.43 to 0.48. Genome wide association studies is

being carried out to identify the genomic regions having significant association with heading date, and the results and biological insights will be presented.

References:

de los Campos G., et al. (2013) *PLoS Genetics* 9: e1003608.

Endelman J. B. (2011) *Plant Genome* 4: 250-255.

Habier D., et al. (2007) *Genetics* 177: 2389-2397.

Kantarski T. et al. (2017) *Theor Applied Genet* 130:137-150.

Meuwissen T., et al. (2001) *Genetics* 157:1819–1829

Pritchard J.K., et al. (2000) *Am J Hum Genet* 67:170-181

Torkamaneh D., et al. (2017) *BMC Bioinformatics*, Article number: 5

PO0511: Other Plant Species

Genotyping-By-Sequencing and Its Applications in Asparagus Breeding Program

Gurleen Sidhu, UNIVERSITY OF GUELPH, Guelph, ON, Canada

The development of new hybrids with improved yield and other economically valuable traits such as quality, disease, replant resistance and longevity is the ultimate objective of the asparagus breeding programs across globe. Understanding the genetic basis of these as well as those that support the development of hybrids, such as tissue culture response and production of berries on male plants (andromonoecy) would benefit a breeding program. However, asparagus being perennial and dioecious, mapping and study of genetic architecture is difficult. Therefore, introduction of genotyping-by-sequencing (GBS) approach is important for overcoming challenges in asparagus mapping providing a method for genotyping thousands of new markers in this crop. GBS is a simple highly multiplexed system for constructing libraries in a range of plant species including those with complex genomes such as asparagus. This technique is becoming an increasingly important technology as the sequencing cost per genotype is quite low and methods are also far less complex compared to other restriction-site associated DNA markers. The technological advancement available through GBS technology provides a new opportunity to expand mapping and studies of genetic architecture in asparagus for many traits. It also allows diversity analysis at a depth not possible with previous methods. Therefore, application of genotyping-by-sequencing (GBS) approach in asparagus breeding program will significantly speed up development of new hybrids and enhance asparagus breeding program in future.

PE0512: Other Plant Species

Developing a Mutant Quinoa Population

Bronte Elizabeth Miller, Brigham Young University, Provo, UT

Quinoa is a grain commonly grown in Peru and Bolivia in South America, but many quinoa varieties grown in other regions around the world are not as successful. Traits such as low heat tolerance, downy mildew susceptibility, low harvest index, and high seed-saponin levels are aspects that need improvement in order to further the production of quinoa outside of its natural environment. This project focuses on understanding the effects of mutagenized genes on quinoa phenotypes. Following a forward genetics procedure, quinoa seeds were soaked in 2% EMS, a mutagen known for inducing point mutations, before being planted. Further generations were bred through self-fertilization, and new phenotypes such as variegated leaves and more complex branching patterns were later observed. DNA from families of interest was analyzed by sequencing and identifying novel SNPs. The expected EMS mutation rate was recorded, and various quinoa families presented mutations that may have had a significant effect on their phenotypes.

PO0513: Other Plant Species

Canada Fleabane Genome Sequence

Martin Laforest, Agriculture and AgriFood Canada, Saint-Jean-sur-Richelieu, QC, Canada, Eric Page, AAC-AAFC, Harrow, ON, Canada and Sara Martin, AAC-AAFC, Ottawa, ON, Canada

Canada fleabane is one of the most economically important weed species worldwide and infestation of herbicide resistant biotypes can result in significant yield losses across a range of cropping systems. As such, Canada fleabane has been the subject of many studies to understand how it has become resistant to herbicides. These studies have not yet identified a mechanism that explains non-target site glyphosate metabolic resistance. We have provided the first report of a chromosome-scale genome sequence for Canada fleabane. Third generation sequencing technology was used to create a genome assembly of 426 megabases, of which 9 chromosome-scale scaffolds cover more than 98% of the entire assembled sequence. This provides the information necessary to allow for the use of powerful genetic tools to detect and map the genes responsible for herbicide resistance. The knowledge gained with the aid of this new tool will be useful to create genetic tests for early diagnostic of resistance and, eventually, for control of this problematic weed. Additionally, the genome sequence will be a resource for studying genetic traits in asters, a large family that represents 10% of the diversity of flowering plants.

PE0514: Other Plant Species

***C. berlandieri* Genome**

Thomas Jordan Kerby, Brigham Young University, Provo, UT

Chenopodium berlandieri is a close relative of *C. quinoa* and is found throughout North America. *C. berlandieri* has adapted to biotic and abiotic conditions that adversely affect quinoa growth, making it an ideal target for the identification of genetic variation that can be used to improve the growth of quinoa outside its native habitat. Here we report the first high-quality reference genome assembly and annotation for *C. berlandieri* *supsp. nuttaliae*, a cultivated Mexican variety commonly known as huauzontle. The assembly was produced using PacBio data and was polished using Arrow and Pilon. The assembly was scaffolded using in vivo Hi-C into 1,194 scaffolds spanning 1.3 Gb. The scaffold N50 is 4.9 Mb. 19.28% of the assembly is in the largest 18 scaffolds, corresponding to the haploid number of chromosomes in *C. berlandieri*. We report genome annotation using RNA-seq data from ### tissues and make a genomic comparison between the genomes of *C. berlandieri* and quinoa. This comparison sheds light on potential genetic variation in *Chenopodium* species that can be used to improve its capacity to be cultivated in lower elevations and higher temperatures.

PO0515: Other Plant Species

Whole Genome Assembly and Characterization of Medicinal Plant *Sophora Flavescens*

Amit Rai¹, Yohei Shimizu², Hideki Hirakawa³, Mami Yamazaki¹ and Kazuki Saito¹, (1)Graduate School of Pharmaceutical Sciences, Chiba University, Chiba, Japan, (2)Chiba University, Japan, (3)Kazusa DNA Research Institute, Kisarazu, Japan

Recent advances in sequencing long reads and computational tools to interpret and assemble raw-reads from multiple sequencing platforms have allowed us to achieve chromosome-scale genome assemblies. Such resources are suitable to explore the evolution of specialized metabolite biosynthesis in plants. Here, we present our progress on whole-genome assembly and comparative genome analysis *Sophora flavescens*. *S. flavescens* is an important medicinal plant widely used in traditional Japanese and Chinese medicine, often in combination with licorice. A mix of *S. flavescens* and licorice (*G. glabra*) tissues have shown to provide improved liver protective and anti-hepatocarcinogenic effects than licorice or Sophora alone. Kmer analysis and flow-cytometry suggested genome size for *S. flavescens* as 1.96Mb. We acquired 80x genome coverage for *S. flavescens* using PacBio-based sequencing. Preliminary assembly using PacBio reads only resulted in entire genome assembly within 473 contigs with contig N50 as 3.76Mb. Using a hybrid assembly approach through a combination of long-reads, short-reads, mate-pair reads, and Hi-C library sequencing, we hope to achieve a high-quality contiguous genome assembly to facilitate the identification, isolation, and editing of useful genes involved in specialized metabolites biosynthesis. Comparative genome analysis of *G. uralensis*, *G. glabra*, *G. inflata*, and *Sophora flavescens* will shed new light on the evolutionary basis of unique chemo-type and pheno-type properties of these plants. There remains much to learn

about the immense diversity of plant metabolism, and this research will contribute to further progress to understand the evolution and diversification of specialized metabolites in planta.

PE0516: Other Plant Species

Developing *Chenopodium ficifolium* As a Diploid Model System Relevant to the Genetic Characterization and Improvement of Allotetraploid *Chenopodium quinoa* (Quinoa)

Madhav Subedi and Thomas M. Davis, University of New Hampshire, Durham, NH

Chenopodium quinoa is an allotetraploid ($2n=4x=36$) species of genome composition AABB, and thus has two distinct diploid ancestors. As has been done for many polyploid crop species, we are developing diploid model systems for genetic studies relevant to quinoa gene identification and marker-assisted breeding. Here we report on our foundational genetic studies in the BB diploid, *Chenopodium ficifolium*. *C. ficifolium* accessions P and QC were collected from weedy locations in Northern New England, and were reciprocally crossed to generate F1 hybrids, which upon molecular confirmation of hybridity, were selfed to produce F2 generation populations. Phenotypic data were collected and F2 generation segregation was evident for several agronomically relevant traits, including flowering time, plant height, internode length, number of branches, branch angle, and chlorophyll content. The *FLOWERING LOCUS T-LIKE* (*FTL*) marker locus was used for genotyping the F1 and the F2 population, and both *FTL* homologs *FTL1* and *FTL2* were characterized molecularly by sequencing the PCR amplicons generated from CrFT345for (5'-GGTTGGTGACTGATATTCCAG-3') and CrFT501Rev (5'-CGCCACCCTGGTGCATACAC-3') primers described by Storchova et al (2015). Marker-trait associations were detected and will be described. The amplicons of trait-associated alleles were sequenced and compared to the quinoa reference genome (Jarvis et al., 2017) to determine the molecular basis for the observed gel mobility polymorphisms, as will be described.

PO0517: Other Plant Species

Genetic Characterization of Two Tetraploid Rose Bi-Parental Mapping Populations

Jeekin Lau, Ellen Young, Natalie Patterson, Patricia E. Klein, **Oscar Riera-Lizarazu** and David Byrne, Department of Horticultural Sciences, Texas A&M University, College Station, TX

Garden roses (*Rosa* spp.) are important ornamental plants in which increased disease resistance and flower productivity are highly desirable. Many rose cultivars are autotetraploid and genetically more complex than their diploid counterparts thus requiring the use of newly released software for analysis. To dissect the genetics controlling disease resistance and flower productivity, two bi-parental populations were developed in 2016. The populations were genotyped using the WagRhSNP 68K SNP array and phenotyped for flower intensity, black spot, cercospora leaf spot, and defoliation. Allele dosage was called using fitPoly and linkage mapping, QTL scans and single marker analysis were done using polypmapR, TetraploidSNPMap, and GAPIT and GWASpoly, respectively. Linkage maps created are similar in length and quality to those in the literature. Single marker analysis and QTL scans found QTL associated with the black spot resistance on chromosomes 3 and 5, with cercospora resistance on chromosomes 1, 3 and 5, defoliation on chromosome 5 and flower intensity on chromosomes 1 and 2.

PE0518: Other Plant Species

Association Mapping of Disease Resistance and Architecture Traits in Diploid Rose Cultivars

Ellen Young, Jeekin Lau, Patricia E. Klein, **Oscar Riera-Lizarazu** and David Byrne, Department of Horticultural Sciences, Texas A&M University, College Station, TX

Roses (*Rosa* spp.) are among the most popular ornamental plants and there is a constant demand for improved cultivars. Of particular interest are cultivars with superior disease resistance, high amounts of flowering, and a more compact, full plant architecture. Some of these traits, however, are time-consuming to phenotype, and breeding would benefit from molecular markers. To that end, 73 diploid rose cultivars were phenotyped for one year in College Station, TX for nine architecture traits (number of primary shoots, plant height, plant length, plant width, longest dimension, plant volume, apical dominance index, growth habit, and growth type), black spot resistance, cercospora resistance, and flowering behavior. Cultivars were also genotyped for single nucleotide polymorphisms via genotyping-by-sequencing. Population structure was investigated with STRUCTURE and ADMIXTURE and

association mapping was performed in GAPIT. Architecture traits were found to have low to moderately high heritability whereas disease and flowering traits had moderate to high heritability. The cultivars were found to comprise five to six subpopulations, and this was used to inform the association mapping. A significant marker for apical dominance index was found on chromosome 2, suggesting that marker-assisted selection for this trait may be feasible. These same traits are currently being analyzed via a QTL analysis in inter-related biparental populations.

PO0519: Other Plant Species

Progress in Genome Sequencing of *Scoliopus bigelovii* (Liliaceae)

Andrew J Marten and **Chris Baysdorfer**, Department of Biological Sciences, California State University, East Bay, Hayward, CA

The genome of *Scoliopus bigelovii* (Liliaceae) is abnormally large (9GB) compared to its nearest relatives (3-4GB). This is presumably due to a disproportionate population of retrotransposons. The morphology and reproductive traits of *S. bigelovii* are also quite different from closely related species. Characterizing the genome of *S. bigelovii* and selected relatives may give valuable insight into the genetic changes responsible for the changes in morphology and reproduction, as well as the potential involvement of retrotransposon activity. However, large genomes with a high level of interspersed repetitive elements pose several challenges to *de novo* assemblies, especially when using short NGS reads. With the advent of long-read sequencing technologies, it is now possible to obtain reads much larger than the repetitive elements. The methods of extracting DNA for sequencing have thus had to adapt to this read-length capacity. A number of techniques for extracting high molecular weight (HMW) DNA have been developed in recent years, but their application to plant species has been relatively limited and is complicated by unique qualities of plant tissues. Furthermore, these techniques often produce some quantity of short fragments alongside the DNA of a desirable length, which leads to a disproportionate output of short reads. In this study, we compared three methods of extracting HMW DNA and three methods of removing fragments that are smaller than typical repeats and the results of these experiments will be presented. In addition, the short and long read sequencing progress for *S. bigelovii* will be discussed.

PE0520: Other Plant Species

Lily Chitinase Gene Family Associated with Resistance to Botrytis Infection

Geung-Joo Lee, Shipra Kumari and Bashistha Kanth, Chungnam National University, Daejeon, South Korea

Lilium, belonging to *Liliaceae* family, is a most important ornamental bulbous, monocot plant with incomparable beauty and commercial importance. *Lilium* has a genome size of about 36GB in total. *Lilium* is threatened by several diseases, caused by fungi, bacteria, viruses, nematodes, and physiological disorders affecting its production and bulb-quality. Upon analysis of RNA-Seq data, we found eight chitinase genes which were differentially expressed among genotypes with variable responses to *Botrytis elliptica*. A phylogenetic analysis showed that they belonged to class I- *LICHI* and *LICHL1A*, V- *LICH2*, *LICH2a*, *LICH2b* & *LICH2c* and VII- *CTL1* and *LICTL1*. After *in vitro* gene expression analysis, *LICHI* gene was highly expressed and type VII *CTL1* and *LICTL1* genes were least expressed. In *in vivo* analysis, *LICHI* and *LICH2b* were highly expressed. After sequence analysis from NCBI, *LICHI* gene showed 70% identity with *RCH10* gene from rice. *LICHI* from *Lilium longiflorum* was cloned in pETDuet vector and the recombinant pETD-LICHI was transformed in *E. coli* codon plus BL21 (DE3) RIL. Total crude protein of *LICHI* was detected as 35.5 kDa size at 37°C under 1 mM IPTG induction, explaining that the *LICHI* gene in lily is successfully translated with *Botrytis* infection.

PO0521: Other Plant Species

The Phylogenetic Context of Herbicide Resistance in *Amaranthus*

Sandra L. Hoffberg and Deren A. Eaton, Columbia University, New York, NY

Modern agriculture practices that rely heavily on herbicides can promote rapid evolution of herbicide resistance in weedy plant species. In recent decades many species in the genus *Amaranthus* (pigweed, amaranth, or waterhemp) have evolved resistance to multiple common herbicides and currently cause significant agricultural losses in the U.S. each year. The history of how these adaptations arose in multiple species provides a powerful way to study

adaptation on large geographic and temporal scales. We compared contemporary samples from across the U.S. and accessions from the National Plant Germplasm System which were collected before the widespread use of herbicides led to strong selection for herbicide resistance at genes that confer resistance to different herbicides. Using genomic data resulting from Restriction enzyme Associated DNA sequencing (RADseq) libraries, we reconstructed a phylogeny including 34 *Amaranthus* species to study the spread of resistance via hybridization and introgression. We found evidence for introgression at the gene that confers resistance to commonly used photosystem II-inhibiting herbicides, but not at genes that confer resistance to other herbicides. Understanding the importance of hybridization in the spread of herbicide resistance in this group of plants can help predict the future spread of herbicide resistance and encourage farmers to use more sustainable weed management that does not promote the evolution of herbicide resistance.

PE0522: Other Plant Species

Deciphering/Comparing Genomes of a Few Landraces of Grain Amaranths from India

Subhashini Srinivasan, Institute of Bioinformatics and Applied Biotechnology, Bangalore, Karnataka, India, Saptarathi Deb, Institute of Bioinformatics and Applied Biotechnology, Bangalore, India and Samathmika Ravi, University of Padova, Legnaro (PD), Italy

Grain amaranths are species producing nutritious grains with especially high lysine, which is one of the major limiting amino acid among the malnourished. India is one of the countries that has enjoyed cultivation of these species for more than a few centuries since the Columbian Exchange and since a ban was imposed on cultivation of these crops in the West. The cultivation of all three grain amaranths species in India for centuries has resulted in many landraces that are optimized for high yield and diverse environmental conditions. It is of interest to study their genomes and compare with other accessions in order to introduce agronomic traits via breeding. Here, we present the genomes of a few landraces and compare them to existing accessions.

PO0523: Other Plant Species

Genome-Wide Transcriptional Profiling to Unravel Self and Non-Self Recognition System Facilitating Self-Incompatibility in Tea [*Camellia sinensis* (L.) O. Kuntze]

Romit Seth^{1,2}, Abhishek Bhandawat², Rajni Parmar², Kamal Dev Sharma¹ and Ram K. Sharma^{*2}, (1)CSK Himachal Pradesh Agricultural University, Palampur, India, (2)CSIR-Institute of Himalayan Bioresource Technology, Palampur (Himachal Pradesh), India

Tea is the most extensively consumed non-alcoholic beverage across the globe known for its wide medicinal properties. However, the major bottleneck lies in maintaining the quality cultivars due to its profuse heterogeneous nature, which is attributed *via* obligate self-incompatibility (SI). Hence, the conventional breeding efforts were preferred over the natural propagation, which has its own disadvantages. The current study unravels the molecular-insights of SI and cross-compatibility (CC) during self (SP) and cross-pollinated (CP) conditions. The cytological investigations revealed anomalous pollen-tube behavior (cessation/deviation) in SP, while, successful fertilization in CP at 48 hours after pollination. The genome-wide transcriptome profiling assisted us in identification of 195 significant differentially expressed genes, subsequently categorized into five phases based on their role in self/non-self pollen-pistil cross-talk. The presence of 182 genes in major hub of predicted protein-protein interactome (PPI) network revealed a complex signalling mechanism during self and non-self recognition system. Moreover, correlating the cytological investigations with tissue-specific (stigma-style & ovary) and event specific (SP & CP) relative expression analysis of 42 key genes suggests existence of pre-zygotic late-acting gametophytic self-incompatibility (LSI) in tea, initiated in styles and sustained up to ovary. These include active involvement of *csRNS*, *SRKs* & *SKIPs* during SP, while, cysteine rich proteins (*RALF*), receptor like kinases (*FER-rlk*, *ANXUR-rlk*), *GPI-Anchored proteins* (*COBL10*, *LLG*) & transcription factors (*MAPK*) during CP probably facilitating fertilization. The current study comprehensively unravels molecular insights of phase-specific pollen-pistil interaction during SI and fertilization, which can be utilized to enhance breeding efficiency and genetic improvement in tea.

PE0524: Other Plant Species

Correlation of Asexual Reproductive Methods and Genome Size within the Genus of Kalanchoe

Bharti Parihar and Ana Almeida, California State University East Bay, Hayward, CA

Kalanchoe Adanson consists of approximately 140 species distributed throughout Madagascar, South and East Africa, and Southeast Asia. This genus is roughly divided into three taxonomic sections, based on the manner through which the species reproduce. Section I includes most of the species that propagate exclusively through sexual reproduction. Section II comprises species that reproduce either via seed production or vegetative propagation. Section III, and the most common strategy implemented by this genus, are those which only reproduce asexually. Asexual reproduction in *Kalanchoe* is commonly a result of plantlet development in the margin of cladodes as in the case of *K. marnieriana*, or on the leaf pedestal with *K. daigremontiana*. We will investigate the correlation between asexual reproduction of *Kalanchoe* and genome size by way of flow cytometry, as this method has contributed immensely towards the understanding of genome size in plants and animals. Previous research reported *K. fedtschenkoi*, another asexually reproducing species, with a genome size of ~260 Mb. Further studies revealed two WGDs (whole genome duplications) in *K. fedtschenkoi*'s lineage, with both duplications occurring close in time. Using *K. fedtschenkoi* as a reference point, it is expected to see a correlation between *Kalanchoe* species which fall into their respective categories of reproduction and genome size. Section I plants are expected to have considerably higher genome sizes, whereas Section III plants will have the lowest. FLOW analysis of additional species serves to provide clarity of phylogenetic relationships between *Kalanchoe* species contributing towards the broader scope of WGDs of land plants.

PO0525: Other Plant Species

A Whole Genome Association Study in *Salvia miltiorrhiza*

Mei Jiang and Chang Liu, Institute of Medicinal Plant Development, Chinese Academy of Medical Sciences&Peking Union Medical College, Beijing, China

Salvia miltiorrhiza is one of the most commonly used traditional Chinese medicine and recognized as a model medicinal plant, which has remarkable curative effect in the treatment of cardiovascular and cerebrovascular diseases and anti-oxidation. Plants produce a variety of metabolites having a critical role in treating disease due to the genomic variation. Here, we produced a whole genome association study of 383 *S. miltiorrhiza* accessions based on the metabolome. This project was composed of five parts. First, these diverse accessions were collected all over China and grown in the same place, each accession obtained multiple plants by root propagation. We collected leaves to extract DNA and root from three different plants per accession to detect metabolome. Second, we resequenced the whole genome of 383 *S. miltiorrhiza* accessions and obtained average 5 GB raw data for each accession, which comprised about 10-fold coverage of the *S. miltiorrhiza* genome. A total of about 8 million SNPs were obtained by both SAMtools and Genome Analysis Toolkit. The genetic structure of the 383 accessions was analyzed through the population structure and principal component analysis, showing these accessions were grouped into two distinct groups. Third, we obtained the metabolic profile of *S. miltiorrhiza* by using a liquid chromatography-tandem mass spectrometry (LC-MS/MS) based method. Hundreds of metabolites were detected, containing the tanshinone, salvianolic acids and compounds in their biosynthesis pathways which are the primarily medicinal constituents. In addition, the population structure was analyzed based on the metabolome. Fourth, a metabolic genome-wide association study was conducted, which obtained hundreds of common variants influencing numerous metabolites. To identify candidate genes related to tanshinone biosynthesis that have not been identified previously, we will look for the protein cluster that is related to the associated metabolic trait encoded at these loci and perform clusters analysis of candidate gene relative to homologous genes with known function. Finally, validating candidate genes by detecting the metabolites of plants were overexpressed and inhibited candidate genes. Our study provides insights into the genetic base of *S. miltiorrhiza* metabolome variation and can facilitate the analysis of the tanshinone biosynthesis pathway and the selection of elite traits of *S. miltiorrhiza*.

PE0526: Other Plant Species

Mapping of Sweet Basil Quality Traits to the Recently Assembled Reference Genome Enabled the Detection of the Plausible Causative Genes

Itay Gonda¹, Renana Milavski¹, Chen Adler¹, Adi Doron-Faigenboim², Ofir Tal¹, Alona Shachter¹, David Chaimovitch¹ and Nativ Dudai¹, (1)Agricultural Research Organization, Ramat Yishay, Israel, (2)Department of Vegetable and Field Crops, ARO, Rishon LeZion, Israel

Sweet basil, *Ocimum basilicum* L., is a well-known herb used in various cuisines, in traditional medicine, as a source for essential oils and as an ornamental. The lack of a reference genome for sweet basil caused a bottleneck in research and breeding. We used our recently assembled genome, comprised of 12,212 scaffolds ($N_{50} > 19\text{Mbp}$), to map quantitative trait loci for various quality parameters. We have genotyped-by-sequencing an F_2 mapping population derived from a cross between the cultivar 'Perrie', a Genovese type green-flowers basil with classic basil aroma, and the cultivar 'Cardinal', a Thai type red-flowers basil with anise-like aroma notes. We have mapped quantitative trait loci for anthocyanins accumulation and aroma volatiles production as well as for fusarium wilt resistance. Candidate gene approach enabled us to suggest the causative genes distinguishing between red anise-like basil to green Genovese basil. We have also detected the plausible causative gene for the resistance of sweet basil varieties to fusarium wilt disease. Moreover, the results show that the two basil subgenomes, originated from the tetraploidy nature of *O. basilicum*, contributed differentially for each trait. This work demonstrates our ability to use the new reference genome to resolve essential traits in sweet basil. That will assist in developing precise molecular markers for traits such as disease resistance and tolerance to environmental stresses and essentially accelerate breeding programs. We will gain a better understanding of the underlying mechanisms and regulations of important metabolic and physiologic processes in the context of polyploid genome structure.

PO0527: Other Plant Species

Molecular Characterization of White River Beardtongue, *Penstemon scariosus* var. *albifluvis*

Sydney J Porter¹, Robert L Johnson¹, Jason M Stettler¹, Mikel R. Stevens¹, Lindsey Meservey¹, Chris D Anderson¹, Robert L Johnson¹, Leigh A Johnson¹, Matthew D Robbins² and Nathan Ricks¹, (1)Brigham Young University, Provo, UT, (2)USDA-ARS Forge and Range Research Laboratory, Logan, UT

Penstemon scariosus is a perennial plant species which thrives in the arid Uinta Basin region of the northern Colorado Plateau. This species is currently divided into four varieties, *P. scariosus* var. *albifluvis*, *P. scariosus* var. *cyanomontanus*, *P. scariosus* var. *garrettii*, and *P. scariosus* var. *scariosus*. Due to habitat destruction and fragmentation from oil and gas development and exploration, *P. scariosus* var. *albifluvis* is being considered for listing under the Endangered Species Act of 1973. At one point *P. scariosus* var. *albifluvis* was recognized as a distinct species but was reclassified to a variety of *P. scariosus* based on similarities of plant morphology. We hypothesize that *P. scariosus* var. *albifluvis* is reproductively isolated and genetically distinct from *P. scariosus*. To test this hypothesis, we collected tissue samples from 66 populations of currently recognized *P. scariosus* varieties as well as four populations of *P. subglaber* as an outlier. We developed ten microsatellite markers for *P. scariosus* and tested the allelic variation between these taxa. Microsatellite data was analyzed using the program STRUCTURE to determine population structure. We found evidence that *P. scariosus* var. *albifluvis* is genetically distinct from all other taxa and should be considered for reclassification as a species. We also found another genetically distinct taxa within the *P. scariosus* collections that is currently undescribed and found morphological and molecular evidence that *P. scariosus* var. *cyanomontanus* and *P. scariosus* var. *garrettii* are indistinguishable.

PE0528: Other Plant Species

Speak to Me Beardtongue: Hierarchical Relationships Among the Six *Penstemon* Subgenera Using Whole Chloroplast Genome Sequences

Jason M Stettler, Lindsey Meservey, Wes Crump, Jed Grow, Nathan Ricks and Mikel R. Stevens, Brigham Young University, Provo, UT

The beardtongues (*Penstemon* Mitchell) are a genus of over 280 species of perennial forbs and sub-shrubs native to North America divided into six subgenera primarily based on floral characteristics. Although this genus has a recent geologic origin, about 2.5 MYA during the early Pleistocene, it has undergone rapid diversification with species occupying all North American habitats including the Arctic Circle, the Central American, alpine meadows, arid deserts, and grasslands. Due to this rapid rate of diversification and speciation observed, many phylogenetic studies using individual or concatenated chloroplast gene sequences have failed to resolve polytomic clades of closely related species within subgenera. We sequenced the complete chloroplast genomes (plastomes) of representative

species from each *Penstemon* subgenera using the Illumina Hi-Seq (2x250) platform, assembled, annotated, and analyzed the plastomes' gene content, gene order and synteny, and Maximum Likelihood phylogenetic relationships among the Lamiales order. We compared whole plastome phylogenies to *matK* and *rbcL* based phylogenies that are commonly used in phylogenetic studies of this group. We found that the whole plastome phylogenies within *Penstemon* genus and within the Lamiales order were different from single chloroplast genes. In addition, the whole plastome based phylogeny has high nodal support which suggests that it provides greater accuracy in describing the hierarchical relationships among taxa than other methods.

PO0529: Other Plant Species

Genomic Consequences of Hybridization in Neotropical *Costus*

Nikolaus Grunder¹, Ana Almeida¹ and Chelsea D Specht², (1)California State University East Bay, Hayward, CA, (2)Cornell University, Ithaca, NY

The evolutionary role of hybridization has been studied extensively for many years. With increased use of molecular tools and genomic analysis, biologists are investigating how hybridization contributes to biodiversity and speciation. It has been hypothesized that hybridization promotes phenotypic diversity and adaptation by allowing for gene flow between distinct evolutionary lineages. Hybridization events frequently result in the formation of allopolyploids, organisms that contain multiple sets of chromosomes derived from different species. Research suggests angiosperms are particularly prone to allopolyploid events, however exactly how allopolyploidy contributes to speciation remains mysterious. I plan to use flow cytometry and cutting-edge genomic sequencing techniques to investigate the outcome of a newly formed hybrid suspected to be the result of a cross between two distinct tropical flowering plant species belonging to the genus *Costus*. The objective of this research is to investigate the genomic consequences of hybridization in terms of structural and functional genomics. More specifically, I intend to study the consequences of hybridization to genomic content and parental gene usage during flower development.

PE0530: Other Plant Species

The Genetics of Planting Density-Dependent Branching in *Chrysanthemum*

Wei Sun, other, Nanjing, China

The architecture of chrysanthemum plants raised for cut flowers is strongly influenced by their ability to form branches. The genetic basis of branching was revealed through a quantitative trait locus (QTL) analysis of an F1 population generated by crossing the varieties 'Nannong Xuefeng' and 'Monalisa' grown under contrasting planting densities. Under the low planting density regime (E1), 12 additive QTLs involving seven branching-associated traits were detected, while under the high planting density regime (E2), the number of QTL detected was only eight. One of the individual QTL accounted for over 10% of the phenotypic variance. Of the 20 QTLs, only one was expressed under both high and low planting densities. A joint QTL analysis across the two environments identified two QTLs which were separately detected in E1. A set of four QTLs exhibiting additive × additive epistasis was identified, few of which had interaction with environments.

PO0531: Other Plant Species

The CmTCP20 Gene Regulates Petal Elongation Growth in *Chrysanthemum morifolium*

Yunxiao Guan, College of Horticulture, Nanjing Agricultural University, Nanjing, China

Chrysanthemum morifolium is one of the most popular ornamental species worldwide, with high ornamental and economic value. Petal size is an important factor that influences the ornamental value. *CmTCP20* is a member of *TEOSINTE BRANCHED1/CYCLOIDEA/PROLIFERATING CELL FACTORS* (TCPs) gene family, which is closely associated with the growth and development of plants. Our previous study found that the expression of *CmTCP20* was obviously down-regulated during chrysanthemum petal elongation, but its function in petal elongation has not yet been revealed. We show here that the overexpression *CmTCP20* in *Arabidopsis* and chrysanthemum leads to similar phenotypes, including larger flower buds (or inflorescences) and longer petals. Interestingly, ectopic expression in *Schizosaccharomyces pombe* yeast cells showed that *CmTCP20* could repress cell division and promote cell elongation. Moreover, the yeast two-hybrid, BiFC and pull-down experimental results indicated that *CmTCP20* may regulate petal size via interacting with CmJAZ1-like and inducing downregulation of *CmBPE2* gene

expression. This study preliminarily clarifies the function of *CmTCP20* on chrysanthemum petal elongation, providing the basic theory for improving the ornamental characteristic of chrysanthemum.

PE0532: Other Plant Species

Understanding the Polyploidisation Process of Hexaploid Chrysanthemum

Natascha H. M. van Lieshout, Martijn van Kaauwen, Linda Kodde, Paul Arens, Marinus J. M. Smulders, Richard G.F. Visser and Richard Finkers, Wageningen University & Research, Plant Breeding, Wageningen, Netherlands

To shed light on the polyploidisation process of cultivated chrysanthemum we develop a high quality genome sequence of a diploid wild species as well as re-sequencing a selection of other possible progenitors of the current, economically important, hexaploid varieties. With a large and highly heterozygous genome even the diploids have to date proven difficult to assemble. Here we present a novel assembly based on long and short read sequencing technology and using polyploid aware assembly tools. We also provide some insight into how the species fit together into the larger chrysanthemum family.

PO0533: Other Plant Species

Progeny Segregation Ratios from RADseq Identify Allohexaploid Inheritance in Eurasian Watermilfoil

Jeff Pashnick, Montana State University, Bozeman, MT

Polyploidy is a common phenomenon in plants. There are two different types of polyploidy – allopolyploidy where chromosomes are inherited disomically, and autopolyploidy where chromosomes are inherited polysomically. While many polyploid plants with high economic value such as wheat (*Triticum aestivum*) and potato (*Solanum tuberosum*) have their chromosomal inheritance type determined, not all systems have the same level of genetic information. Without information on polyploid chromosomal inheritance, studies linking genotype to phenotype (such as quantitative trait locus (QTL) mapping) are not possible. Eurasian watermilfoil (*Myriophyllum spicatum*, *sensu lato*) is hexaploid and one of the most heavily managed aquatic weeds in the United States. Recently, herbicide resistance has been detected in Eurasian watermilfoil, sparking interest in determining the genetic architecture of those traits. In this study we selfed a genotype of 2,4-D resistant watermilfoil to determine the chromosomal inheritance pattern. RADseq data were collected on this population providing 293 molecular markers to determine their most likely chromosomal inheritance with the R package PolyRAD. We found that all 293 molecular markers had higher likelihood of being inherited as an allohexaploid (disomic). No molecular markers in this data set were more likely to be inherited under the autohexaploid model (polysomically). Due to the evidence provided in this study we believe that Eurasian watermilfoil is an allohexaploid, enabling molecular markers to be treated as a diploid for the future construction of a linkage map and QTL analysis for 2,4-D response.

PE0534: Other Plant Species

The Chromosome-Scale Reference Genome of Black Pepper Provides Insight into Piperine Biosynthesis

Zhongping Xu and Shuangxia Jin, National Key Laboratory of Crop Genetic Improvement, Huazhong Agricultural University, Wuhan, Hubei, China

Black pepper (*Piper nigrum*, $2n = 52$), dubbed the ‘King of Spices’ and ‘Black Gold’, is one of the most widely used spices. It has provided a source of food spice and phytomedicines for thousands of years due to its characteristic flavor, aroma and the pharmaceutical effects of piperine. *Piper* is also important for evolutionary studies because it is included among the strikingly diverse and long-isolated lineages of basal angiosperms. Here, we present a high-quality reference genome assembly by integrating PacBio Sequel, 10x Chromium, BioNano DLS optical mapping, and Hi-C mapping technologies. The 761.2 Mb sequences (45 scaffolds with an N50 of 29.8 Mb) are assembled into 26 pseudochromosomes, representing the first sequence of a species in the Piperales. A phylogenomic analysis of representative plant genomes places magnoliids as sister to the monocots-eudicots clade and indicate that black pepper has diverged from the shared Laurales-Magnoliales lineage approximately 180 million years ago.

Comparative genomic analyses reveal specific gene expansions in the glycosyltransferases (*GTF*), cytochrome P450 (*CYP*), shikimate hydroxycinnamoyl transferase (*HCT*), lysine decarboxylase (*LDC*), BAHD acyltransferases (*BAHD-AT*) and SCPL acyltransferases (*SCPL-AT*) gene families. Comparative transcriptomic analyses disclose berry-specific upregulated expression in representative genes in each of these gene families. These data provide an evolutionary perspective and shed light on the phenylpropanoid pathway (KEGG PATHWAY: map00400), L-lysine metabolism (KEGG PATHWAY: map01064) and acyltransferase metabolic processes relevant to the molecular basis of species-specific piperine biosynthesis. Sequencing the black pepper genome has advanced our understanding of the unique piperine biochemistry of black pepper. Our study therefore provides valuable insights that may serve as a foundation for future research on Piperales taxonomy and piperine biosynthesis, leading to a better understanding of the evolution, phytochemistry and ecology of the *Piper* genus.

PO0535: Other Plant Species

Unveiling Transcriptome Composition of *Festuca brevipila* through PacBio Isoform Sequencing

Yinjie Qiu, Cory D. Hirsch and Eric Watkins, University of Minnesota, St. Paul, MN

Hard fescue (*Festuca brevipila* Tracey, $2n=6x=42$) is a cool-season turfgrass that has fine leaf texture that does well in low input areas. Breeding and genetics studies of *F. brevipila* have been limited due to the nature of its complex hexaploid genome. To advance our knowledge of *F. brevipila* genomics research, we used PacBio isoform sequencing to develop the reference transcriptome of this taxon. Here, we report the *F. brevipila* reference transcriptome generated from the root, crown, leaf, and seed head tissues using 4 SMRT cells. We obtained 38,595 non-redundant full-length transcripts which had the N50 of 2,585 bp. respectively. Transcriptome annotation was done using NCBI NR protein and Uniprot databases returned with 36,075 (93.47%) and 29,670 (76.88%) transcripts have been annotated, respectively.

In a previous study, we observed enhanced snow mold (*Microdochium nivale*) disease resistance in fungicide (propiconazole) treated plants vs control plants half a year after the last fungicide was applied. Transcriptome analysis suggested 1,849 genes, including 39 P450 genes were expressed differentially at a significant level six hours after snow mold inoculation. These P450 genes identified in this comparison were found to be related to salt tolerance, flavonoid biosynthesis, defense response. When comparing with/without snow mold inoculation within each treatment group: In fungicide-treated plants, we identified 1,140 genes differentially expressed genes, 24 of which were P450 genes; in water-treated plants, we found 2,283 genes that were differentially expressed, 60 of which were P450 genes. Most of the P450 genes in these comparisons were related to fatty acid biosynthesis and the oxidation-reduction process. Overall, the *F. brevipila* reference transcriptome provides a powerful foundation for transcriptome study of this complex taxon. Results from this study will help researchers learn more about the pink snow mold disease resistance mechanisms.

PE0536: Other Plant Species

Annotation of the Mediterranean Tall Fescue (*Festuca arundinacea* Schreb.) Plastid Genome

Md Shofiqul Islam¹, Gretta Leigh Buttelmann², Konstantin Chekhovskiy¹, Taegun Kwon¹, Jennifer Black¹ and Malay C. Saha¹, (1)Noble Research Institute, LLC, Ardmore, OK, (2)Missouri State University, College of Agriculture, Springfield, MO

Mediterranean tall fescue (MTF) is a forage grass species that exhibits incomplete summer dormancy to avoid extreme drought and heat during summer months and starts re-growing under favorable conditions in the fall. Despite its agronomic importance, the genomic constitution of MTF largely remains unknown. Being maternally inherited, we aimed to sequence, assemble, and annotate the chloroplast genome of MTF to develop MTF specific molecular markers as well as to identify its ancestors. Intact chloroplasts were isolated from green leaves not older than eight weeks. A quick and inexpensive protocol was developed to obtain DNA from intact chloroplasts purified through one layer of 30% Percoll. The KAPA HyperPlus Kit (Roche) was used for library preparation using 200 ng of the extracted DNA. Sequencing was performed in Illumina MiSeq next generation sequencer. A total of 4,223,102 paired-end reads were obtained with an average read length of 300 bp. The reads were quality filtered and contaminating mitochondrial reads (0.23%) were removed from the data set by performing a reference assembly against mitochondrial genome of perennial ryegrass (*Lolium perenne* L.). The filtered reads (2,978,067) were *de*

novo assembled using CLC Genomics Workbench resulted in 209 contigs, of which 22 contigs ranging from 1,064 to 8,163 bp were identified as chloroplast sequences. Reference guided assembly identified 8,443 reads totaling of 1,327,800 bases, which covered 96% of the plastid genome sequence of the continental tall fescue with an average read depth of 10-fold. Reference assembly resulted in eight gaps ranging from 0.2 to 2.1 kb between MTF and reference genome. Primers were designed and synthesized to PCR amplify the gaps. We are currently working on the assembly and annotation of the chloroplast genome of MTF. This work will constitute a solid basis to perform comparative analysis of plastid genomes within *Festuca* and between closely related grass species to accelerate marker-assisted tall fescue breeding program for greater persistence and productivity.

PO0537: Other Plant Species

Assembly and Annotation of the Cheatgrass (*Bromus tectorum*) Genome

Craig E. Coleman¹, Daniel Duraes Machado¹, Ian C. Burke², Samuel R. Revolinski² and Peter J. Maughan¹, (1)Brigham Young University, Provo, UT, (2)Washington State University, Pullman, WA

Cheatgrass or downy brome (*Bromus tectorum* L.) is an invasive plant that has infested western North American rangelands. It is also a severe weed in western pasturelands and wheat fields. The extent of the invasion is expansive, occupying millions of hectares while displacing native flora and fauna. In addition to the ecological damage, the role of cheatgrass as a fine fuel for rangeland fires has resulted in large economic losses including property damage, unhealthy air quality due to smoke and fine particulates, and costs associated with fighting fires. Cheatgrass is very successful in areas disturbed by grazing and fire, leading to an expansion of the invaded range that is extremely concerning to land managers. As a weed in cropping systems, cheatgrass can reduce winter wheat yields by as much as 90%. Cheatgrass has also evolved multiple resistance mechanisms to a range of herbicides, including glyphosate. Genetic work on cheatgrass has focused on how different haplotypes are distributed across the landscape and how that distribution may be related to ecological adaptations to varying environments. For instance, there are distinctive genetic and physiological differences between cheatgrass found in mid-elevation sagebrush-steppe zones and cheatgrass found at lower elevations in the Chihuahuan and Sonoran Deserts, or across the climatically variable agroecosystems of the Pacific Northwest. The genetic studies have relied on a small number of markers, including isozymes, microsatellites and single nucleotide polymorphisms with a limited capacity to address important questions about the adaptation of cheatgrass to different environments. We report the completion of a high-quality canu-based assembly of the cheatgrass genome assembled using 83X coverage of PacBio long reads. The genome was polished with two rounds of Arrow, followed by indel correction with ~60X coverage of Illumina reads using Pilon. The canu assembly consists of 447 contigs with an N50 and L50 of 26.9 Mb and 29, respectively and spans 2.48 Gb. Chromatin contact maps (Hi-C; Dovetail Genomics, Inc.) were used to scaffold the canu assembly to chromosome scale, which was then annotated using Repeatmodeler and Maker. To better understand the genetic variation within the species, we have resequenced, at 10X coverage, ten different cheatgrass haplotypes collected from warm deserts, salt deserts, wheat fields, and locations in the central Asian native range. The completed cheatgrass genome will be a valuable resource for studying cheatgrass invasiveness and weediness, including its adaptation to different environments across Western North America, perhaps leading to the development of tools for checking its invasive expansion.

PE0538: Other Plant Species

Tall Fescue Gene Expression in Different Tissues Under Stress Conditions

Randy Dinkins¹, Padmaja Nagabhyru² and Christopher Schardl², (1)USDA-ARS, Lexington, KY, (2)University of Kentucky, Lexington, KY

Tall fescue is a major forage and turf species for the transition zone in the United States. A major factor for this is the ability to survive stress environments. However, its exceptional stress resistance is not entirely a function of the plant, but is augmented by a seed-transmitted fungal endophyte, *Epichloe coenophiala*, although the exact mechanisms responsible for conferring environmental stress tolerance during symbiosis remain largely unknown. Since the endophyte is only vertically transmitted through the seed in tall fescue, propagation and stability of the fungus in the plant is necessary in the crown tissues over the life of the plant. It is likely that metabolic crosstalk between the grass and the fungus results in complex up- and down-regulation of both fungal and plant genes and concentrations of the products they code for in complex ways that contribute to providing for stress tolerance. In order to identify potential regulatory pathways, a transcriptome analysis was conducted in the leaf, pseudostem,

crown and roots of tall fescue plants following abiotic stress in the greenhouse. Comparisons between the fungal gene expression in the pseudostem and crown where the fungus resides, and the four plant tissues was done. Both plant and fungal gene expression profiles demonstrated that genes normally associated with stress responses were differentially expressed when compared to unstressed control plants. The presence of the endophyte may have primed the plant for enhanced stress resistance, but plant genes affected by the endophyte were few. While many pathways affected by the stress were similar across the four tissues, tissue specific differential expression revealed that the leaf was the most affected by the stress treatment, while the crown was the least affected.

PO0539: Other Plant Species

Identification of Genetic Loci Associated with Self-Compatibility, Pre-Harvest Sprouting and Maturity Time Using Targeted Next-Generation Sequencing in Common Buckwheat (*Fagopyrum esculentum*)

Ryoma Takeshima¹, Eri Ogiso-Tanaka¹, Yasuo Yasui² and Katsuhiro Matsui¹, (1)National Agriculture and Food Research Organization, Ibaraki, Japan, (2)Kyoto University, Kyoto, Japan

Common buckwheat (*Fagopyrum esculentum*) is a nutritionally dense crop widely grown in the temperate zone, however, the genetic breeding has not kept pace with other major crops. One problem with genetic breeding is self-incompatibility (SI), which based on the heteromorphic incompatibility system.

We previously developed a self-fertile common buckwheat line as a mid-mother parent. Here, we identified the genetic loci of the self-compatibility (SC) and pre-harvest sprouting (PHS) tolerance, which is an important trait in buckwheat using three F₂ populations (cross A to C) derived from SI × SC lines. These three populations showed different segregation patterns for PHS tolerance suggested the major tolerance genes in cross A are dominant and those in cross B and C are recessive. To develop the makers which tightly link with SC and PHS tolerance, we performed re-sequencing of the bulked DNA which showing opposite trait value in cross A and developed 100 makers as SC linked, 100 makers as PHS linked and 300 makers as the whole-genome region. Using this 500 maker-set, we genotyped the progenies of cross A to C by Ion AmpliSeq technology and constructed genetic linkage maps.

The SC linked makers clustered slightly closet region with the phenotypic maker as SI/SC in all three populations. The PHS linked makers separated four linkage groups. In cross A, two major QTLs for PHS were detected near regions of PHS linked makers. In contrast, four and two QTLs of Cross B and C were detected in different linkage groups from Cross A. On top of that, we also detected some QTLs for maturity time. These are the first results to identify the genetic loci of SC, PHS tolerance and Maturity time. A targeted next-generation sequencing genetic approach is an effective way to develop a maker-set and identify the genetic loci for important traits in buckwheat.

PE0540: Other Plant Species

A Comparison of Leaf and Root Transcriptomes of Three Contrasting Genotypes of *Brachiaria* in Response to Increasing Water Stress

Charlotte M Jones¹, Jose De Vega², Ann Thomas³, Dagmara Gasior⁴, John Harper⁴, Jacobo Arango⁵, Margaret L. Worthington⁶, Ian Armstead³ and **Narcis Fernandez-Fuentes**⁴, (1)IBERS, Aberystwyth University, Aberystwyth, ceredigion, United Kingdom, (2)Earlham Institute, Norwich, United Kingdom, (3)IBERS, Aberystwyth University, Aberystwyth, United Kingdom, (4)Aberystwyth University, Aberystwyth, United Kingdom, (5)International Center for Tropical Agriculture, Cali, Colombia, (6)Department of Horticulture, University of Arkansas, Fayetteville, AR

Brachiaria (Trin.) Griseb (syn. *Urochloa* P. Beauv.) is a C4 grass genus belonging to the Panicoideae. Native to Africa, it is now widely grown as a forage grass in tropical areas worldwide and is the subject of intensive breeding and germplasm selection, particularly in S. America. Tolerance to abiotic stresses such as aluminium and drought are major breeding objectives. In this study we present the transcriptomic profiling of drought stressed leaves and roots of three *Brachiaria* hybrid genotypes selected from CIAT breeding program: Br12/3659-17 (*gt-17*), Br12/2360-9 (*gt-9*) and Br12/3868-18 (*gt-18*); previously characterised as having good, intermediate and poor tolerance to drought, respectively. RNA was extracted from leaf and root tissue of plants established in vermiculite

and watered with nutrient solution at estimated water contents (EWC) of 35, 15 and 5%. RNAseq was carried out using standard protocols. Differentially expressed genes (DEGs) were compared between different EWCs, 35/15, 15/5, and 35/5 using DESeq2. Overall, the proportions of DEGs enriched in all three genotypes varied in a genotype-dependent manner in relation to EWC comparison, with *gt-9* and *gt-18* being more similar to each other than to *gt-17*. More specifically, an interesting contrast was that GO terms relating to carbohydrate and cell wall metabolism in the leaves were enriched by up-regulated DEGs for *gt-9* and *gt-18*, but down-regulated DEGs for *gt-17*. Additionally, across all genotypes, analysis of DEG KEGG enzyme activities indicated an excess of down-, as compared to up-regulated likely apoplastic peroxidases in the roots as water-stress increased. This suggests that changes in root cell-wall architecture may be an important component of the response to water-stress in *Brachiaria*.

PO0541: Other Plant Species

Karyotype and Identification of Each Chromosome of Allotetraploid Grassland Plant *Leymus chinensis*

Lei Wu, Meiping Sun, Na Lyv and Zhiyong Xiong, Inner Mongolia University, Huhhot, China

Leymus chinensis (Trin.) Tzvel ($2n=4x=28$), closely related to Triticeae cereals, is a native perennial rhizomatous grass with high palatability and forage value. It is also considered as one of the most promising grass species for grassland rehabilitation and reconstruction in North China. However, since the large genome size and as a allotetraploid species with unidentified parental genomic origins, the genome composition and the evolutionary history of *Leymus chinensis* are largely unknown. In this study, we performed low coverage sequencing of *Leymus chinensis* and the composition and abundance of repetitive sequences were analyzed. In addition, the distribution patterns and chromosomal localizations of the main ten satellite repeat sequences were characterized by FISH. The fine karyotype was established and each chromosome of *Leymus chinensis* could be identified by using five tandem repeats (CL6, CL121, CL162, 5S rDNA and 26S rDNA). We also connected the cytogenetic map of *Leymus chinensis* with the genetic linkage map of wheat using single copy genes as probes. This robust chromosomal identification system will have biological applications for the understanding of chromosome structure, chromosome pairing, and genome evolution in *Leymus chinensis*.

PE0542: Other Plant Species

Are *Carex albida* L.H. Bailey and *Carex lemmonii* W. Boot ‘Genomic Synonyms’? Genomic Comparisons within *Carex lemmonii* for Conservation Purposes

Austin Quach, Ana Almeida and Alyssa Jenkins, California State University East Bay, Hayward, CA

Carex albida H. L Bailey is a sedge native to California and is endemic to the Pitkin Marsh in Sonoma County, CA. In general, the genus *Carex* plays a crucial role in the ecosystem by maintaining water quality, soil quality, and acts as a food source or natural habitat for wildlife. Listed as an endangered species, this population was recently grouped together with the species *Carex lemmonii* due to a morphological study. With little to no research conducted on the genomic level of this species, the population of *Carex albida* may continue to be synonymous with *Carex lemmonii*. This will impact conservation procedures for a species that was once considered endangered to California since 1979. The objective of this experiment is to obtain various samples of populations of *Carex lemmonii* throughout California and compare the genomic sequence to the population that was previously considered *Carex albida*. In order to execute this, the genome from an individual plant from the Sonoma County population will be sequenced at high coverage (>20x coverage). Other populations of *Carex lemmonii* samples will be sequenced at low coverage as well. The comparison of the obtained sequences will help unravel key differences between the different populations and hopefully help develop methods to better conserve the population of *Carex lemmonii* found in Sonoma County, CA.

PO0543: Other Plant Species

The Amborella Pangenome Suggests Gene Presence/Absence Variation Is Associated with Environmental Adaptation

Haifei Hu¹, Armin Scheben¹, Brent Verpaalen¹, Soodeh Tirnaz¹, Philipp E. Bayer¹, Richard G. J. Hodel², Jacqueline Batley¹, Douglas E. Soltis³, Pamela S. Soltis³ and David Edwards¹, (1)University of Western

Australia, Perth, WA, Australia, (2)University of Michigan, Ann Arbor, MI, (3)University of Florida, Gainesville, FL

Amborella trichopoda (Amborellaceae) is the single living sister species of all other extant flowering plants. *Amborella* occurs only in rain forest habitats on the remote island of New Caledonia, in a patchy distribution that reflects a series of genetic bottlenecks and population expansions over the past million years. The geological history of New Caledonia is unique in that the island was likely submerged for millions of years, re-emerging only 33-38 mya, with the present flora of the island therefore representing recolonisation and evolution since that time. The above features make *Amborella* an important species in which to study genome variation in relation to local environmental adaptation. Genome variation in plants often includes gene presence/absence variants (PAVs), with as many as 51% of genes being dispensable in some species. Here, we assembled the *Amborella* pangenome based on ten individuals representing most of the known natural distribution. The pangenome assembly contains 2,765 additional genes (9.2%) compared with the published reference genome. Dispensable genes are enriched in gene ontology (GO) terms associated with abiotic stress, suggesting that gene content may be driven by environmental adaption to specific regions on the island. Compared with crop pangenomes, the *Amborella* pangenome contains fewer disease resistance genes analogs (RGAs), and the RGAs it does contain are less likely to be dispensable. This pangenome for the sister species to all other extant angiosperms offers a chance to study pangenome structure and content in a non-crop plant species and provides a phylogenetic perspective for studying the evolution of PAVs.

PE0544: Other Plant Species

Chromosome-Level Genome Assembly of Liverwort *Marchantia polymorpha* and Updates of the Genome Database MarpolBase

Yasuhiro Tanizawa¹, Yagura Masaru¹, Mika Sakamoto¹, Takako Mochizuki¹, Sean A. Montgomery², Bence Galik³, Shohei Yamaoka⁴, Ryuichi Nishihama⁴, Katsuyuki T. Yamato⁵, Takayuki Kohchi⁴, Yasukazu Nakamura¹ and Frederic Berger², (1)National Institute of Genetics, Mishima, Japan, (2)Gregor Mendel Institute of Molecular Plant Biology, Vienna, Austria, (3)Vienna Biocenter Core Facilities GmbH (VBCF), Vienna, Austria, (4)Kyoto University, Graduate School of Biostudies, Kyoto, Japan, (5)Kindai University, B.O.S.T., Kinokawa, Japan

The liverwort *Marchantia polymorpha* is a member of a basal land plant lineage and spends most of its life cycle as a haploid gametophyte. Along with its basic characteristics – simple but common with other terrestrial plants, easy observation and manipulation under laboratory conditions – make this organism an attractive model for molecular genetics and plant developmental biology.

We have developed MarpolBase, a genome database for *Marchantia polymorpha*, as a data hub for the draft reference genome ver. 3.1. To promote the use of the annotated genome data, MarpolBase features a genome browser, sequence similarity searches using BLAST and GMAP, and various utility tools.

Recently, we newly obtained a chromosome-level genome assembly by leveraging PacBio long-read sequencing and Hi-C genome scaffolding, which consists of 9 chromosomal sequences and additional 435 unplaced scaffolds representing 215 Mbp and 2.8 Mbp in total, respectively. The new genome sequences, referred to as ver. 5.1, harbor 19,421 predicted protein-coding loci with 24,751 transcript models, which include 673 newly identified and 303 modified transcript models based on manual inspection.

In this poster, we will present gene annotation on the ver. 5.1 genome and the update of MarpolBase to accommodate genomic data for ver. 5.1.

PO0545: Other Plant Species

Preliminary Genome Assemblies of Butterworts (*Pinguicula*, Lentibulariaceae) suggest both Divergent and Convergent Evolution of Genomic Features Important in Prey Digestion

Steven J. Fleck¹, Katie Nolan², Kylie Bocklund², Tianying Lan³, Crystal Tomlin¹, Enrique Ibarra-Laclette⁴, Martín Mata Rosas⁵, Todd Michael⁶, Tanya Renner² and Victor A. Albert¹, (1)University at

Buffalo, Buffalo, NY, (2)Penn State University, State College, PA, (3)University at Buffalo, Amherst, NY, (4)Red de Estudios Moleculares Avanzados, Xalapa, VL, Mexico, (5)INECOL, Xalapa, VL, Mexico, (6)J. Craig Venter Institute, La Jolla, CA

Carnivorous plants provide a unique system for studying convergent evolution. Plant carnivory has evolved independently at least ten times in the flowering plants, with many distantly related carnivorous species sharing traits important for the capture and digestion of prey. Prey digestion is a vital component of the carnivorous syndrome and is accomplished through the production of digestive enzymes, such as proteases, within trap tissues. This preliminary study uses highly contiguous Oxford Nanopore-derived genomes to study expansions of protease genes within the carnivorous butterworts (genus *Pinguicula*). *De novo* assemblies of three butterwort genomes are compared against other carnivorous species for expansions in protease genes known for trap specific expression and use in prey digestion. Genome comparisons are made using the online genome evolution tool, CoGe. We report that while butterworts possess tandemly duplicated cysteine protease genes, they are not multiplied to the extent known in *Utricularia gibba*, another carnivorous species within the same family, the Lentibulariaceae. Additionally, butterwort genomes contain interspersed tandem duplications of serine proteases and acyl hydrolases not known from the *Utricularia gibba* genome. This finding is noteworthy because distantly related carnivorous plants, such as *Nepenthes*, abundantly express serine proteases in their pitcher fluid, suggesting potential convergent evolution in digestive enzyme composition.

PE0546: Other Plant Species

deciphering the High Quality Genome Sequence of Coriander

Xiaoming Song, North China University of Science and Technology, Tangshan, China and Xiyin Wang, North China University of Science and Technology, Watkinsville, GA, China

Coriander (*Coriandrum sativum* L. $2n = 2x = 22$), a plant from the Apiaceae family, also called cilantro or Chinese parsley, is a globally important crop used as vegetable, spice, fragrance, and traditional medicine. Here, we report a high-quality assembly and analysis of its genome sequence, anchored to 11 chromosomes, with total length of 2,118.68 Mb and N50 scaffold length 160.99 Mb. We found that two whole-genome duplication events, respectively dated to ~45-52 and ~54-61 million years ago, were shared by the Apiaceae family after their split from lettuce. Unbalanced gene loss and expression observed between duplicated copies produced by these two events. Gene retention, expression, metabolomics and comparative genomic analyses of Terpene synthase (TPS) gene family, involved in terpenoid biosynthesis pathway contributing to coriander's special flavor, revealed that tandem duplication contributed to coriander TPS gene family expansion, especially compared to their carrot counterparts. Notably, a TPS gene highly expressed in all 4 tissues and 3 development stages studied, is likely a major-effect gene encoding linalool synthase and myrcene synthase. The present genome sequencing, transcriptome, metabolome and comparative genomic efforts provide valuable insights into the genome evolution and spice trait biology of Apiaceae and others related plants, and facilitated further research into important gene functions and crop improvement.

PO0547: Other Plant Species

Whole Genome and Transcriptome Resource for Tall and Dwarf Varieties of Coconut

Cynthia Saloma¹, Ma. Anita M. Bautista¹, Charles Anthon E. Cadorna¹, Ma. Regina G. Punzalan¹, Dianne J. Acoba¹, Ramon L. Rivera², Susan M. Rivera³, Ernesto E. Emmanuel² and Zenaida V. Magbanua¹, (1)Philippine Genome Center, University of the Philippines, Quezon City, Philippines, (2)Philippine Coconut Authority (PCA), Zamboanga City, Philippines, (3)Philippine Coconut Authority, Zamboanga City, Philippines

Coconut, *Cocos nucifera* L., is an important crop in the tropics. Its edible and non-edible by-products include copra, coconut oil, lumber, coco water and coco coir used for making ropes and matting. Several varieties of coconut palms are used in commercial planting. A well annotated transcriptome and whole genome reference sequence for coconut is an important resource to study the biology of this palm species that is cultivated in 93 countries. Coconut can generally be divided into two main varieties, tall and dwarf. In the Philippines, the Laguna Tall (LAGT) cultivar is widely used in breeding studies and was the source of the first parental line for the first hybrid generated with another local cultivar, Catigan green dwarf (CATD). Here, we present a comprehensive *de novo* draft assembly of the coconut genome and transcriptome from both these tall and dwarf varieties. LAGT and CATD were sequenced

using a combination of short (Illumina) and long-read (PacBio) sequencing. Using K-mer analysis and flow cytometry, the estimated genome size of *Cocos nucifera* var CATD is 2.9 Gb and 2.6 Gb for LAGT. Retrotransposons are most abundant in coconut compared to either the date or oil palm and could account for the genome size expansion in coconut. The difference in the genome sizes of the tall and dwarf variety could also largely be accounted for by higher repeat content (80%) in the dwarf variety as compared to the tall variety (70%). Transcriptome from leaves, nuts, flowers of mature coconut were sequenced using Illumina Hiseq2000 and assembled *de novo* using various transcriptome assemblers generating a total of 79,263 transcripts from the combined transcriptomes. The estimated number of unigenes in LAGT is 68,147 and 55,876 for CATD. This resource provides a comprehensive assembly of coconut genome and transcriptome and is useful as a molecular toolkit to investigate biological processes in this plant species. It is also an invaluable resource for molecular assisted breeding in coconut varietal improvement.

PE0548: Other Plant Species

The Genome of Charophycean Green Algae *Zygnema Circumcarinatum* Provides Insights into the Early Evolution of Land Plants

Xuehuan Feng¹, Jinfang Zheng¹ and **Yanbin Yin**², (1)University of Nebraska-Lincoln, (2)University of Nebraska-Lincoln, Lincoln,, NE

The charophycean green algae (CGA) species, *Zygnema circumcarinatum*, represents the closest algal relatives to land plants. Sequencing its genome thus will contribute to the understanding of the origin and early evolution of land plants. The genomes of four *Z. circumcarinatum* strains: hypothetically diploid UTEX 1559 [unfunctional mating +], and haploid UTEX 1560 [mating type -], SAG 698-1a [mating type +] and 1b [mating type -] have been sequenced by our lab. Comparing the *Zygnema* genomes with genomes of other green algae and land plants can elucidate the genomic basis of adaptation to the terrestrial environment.

Using the 3rd generation long read technology, an Oxford Nanopore sequencing of UTEX 1559, SAG 698-1a and 1b were finished. Meanwhile, the 2nd generation Illumina sequencing reads have also been generated for the four *Zygnema* strains. RNA sequencing has also been done for SAG 698-1b with desiccation and cold treatments. A hybrid assembly of the draft genome was finished for SAG 698-1b, which had a genome size of 65 Mb (contig N50=154,887bp), close to the estimated genome size of ~ 64 Mb by a DAPI experiment staining the nuclear DNA, a flow cytometry experiment and a k-mer frequency analysis. Preliminary genome annotation indicated that this genome contained 10.95% repeat elements, 0.56% retro elements, 0.56% LTR elements and 0.26% DNA transposons, respectively. An ab initio gene prediction using the program MAKER with the help of RNA-seq data predicted 17,460 protein coding genes. Other analyses, such as evolution features in terrestrialization, cell wall evolution, phytohormone regulation, are still on going. The complete plastome (157,548 bp) and mitogenome (216,223 bp) of SAG 698-1b were also assembled. In the plastome, 93 protein coding genes and 34 tRNA genes were annotated. Also, 76 protein coding gene and 26 tRNA genes were annotated in the mitogenome.

PO0549: Other Plant Species

Water Stress, Apomixis and Gene Expression in the Diplosporous Grass *Eragrostis Curvula*

Juan Pablo Selva¹, Diego Zappacosta^{1,2}, José Carballo^{1,2}, Juan Manuel Rodrigo^{1,2}, Andres Bellido¹, Cristian Gallo¹, Jimena Gallardo^{1,2} and **Viviana Echenique**^{1,2}, (1)Centro de Recursos Naturales Renovables de la Zona Semiárida (CERZOS) Universidad Nacional del Sur-CONICET, Bahía Blanca, Argentina, (2)Dpto. Agronomía, Universidad Nacional del Sur (UNS), Bahía Blanca, Argentina

Eragrostis curvula has become a model to understand apomictic mechanisms, specially diplosporous development since our group have generated many plant and genomic resources to study this trait, like a mapping population, a reference genome, the first high density linkage map, sRNAs and transcriptomes databases. The characterization of the molecular pathways underlying apomixis is a prerequisite to transfer the trait to economically important crops. In nature it is possible to observe mainly facultative genotypes of this grass, which retain a certain percentage of sexual pistils that is increased by external factors such as stress conditions, indicating that some regulators activated by stress could be present affecting the apomixis/sex switch. A series of experiments with *E. curvula* plants of the facultative apomictic cv. Don Walter under control and water deprivation conditions were performed in order to

associate the increase of sexual embryo sacs in apomictic plants with the differential expressed genes in inflorescences. Three control and three stressed plants were used for the experiments. Inflorescences were collected for cytoembryological analysis and for RNA sequencing (Illumina HiSeq1500 platform). As a result, 501 differentially expressed transcripts between control and stressed plants were found, 305 out of them were annotated and their expression mainly coincided with up or downregulated pathways previously associated with apomixis, like ubiquitin mediated proteolysis, RNA degradation and vesicle trafficking molecular pathways among the most relevant. Differentially expressed genes involved in these pathways were downregulated in stressed plants and its role in apomixis is being considered.

PE0550: Other Plant Species

Novel Genomic Resources for the Polyploid Forage Grass *Eragrostis Curvula*

José Carballo^{1,2}, Cristian Gallo², Juan Pablo Selva², Ingrid Garbus², Diego Zappacosta^{1,2}, Bruno A. Santos³, Emidio Albertini⁴, Vicent Were⁵, Nick Talbot⁵, Matthew James Moscou⁵, Mario Caccamo³ and **Viviana Echenique**^{1,2}, (1)Dpto. Agronomía, Universidad Nacional del Sur (UNS), Bahía Blanca, Argentina, (2)Centro de Recursos Naturales Renovables de la Zona Semiárida (CERZOS) Universidad Nacional del Sur-CONICET, Bahía Blanca, Argentina, (3)NIAB, Cambridge, United Kingdom, (4)University of Perugia, Perugia, Italy, (5)The Sainsbury Laboratory, Norwich, United Kingdom

Eragrostis is a polyphyletic genus with more than 400 species originated from Africa and now distributed in tropical and mid-warm season regions all over the world. Weeping love grass (*E. curvula*) is an apomictic C4 grass from this genus used as forage in marginal areas due to its resistance to biotic and abiotic stresses. The basic chromosome number of this grass is X=10 with ploidy levels ranging from 2x to 8x. *E. curvula* has been classified as an orphan, or underutilized crop and despite its importance, few investigations have been performed on this species.

We have generated sequencing data for genome, transcriptome and miRNA libraries obtained from different accessions. Three genomes have been sequenced, one diploid sexual and two apomictic tetraploids and the first high density linkage map for the species was obtained following a genotyping-by-sequencing (GBS). Functional analysis of the genetic and epigenetic mechanisms underlying apomixis were initiated through the sequencing of diverse miRNA libraries and transcriptomes. Finally, 17 accessions were sequenced through RNA-seq to discover NLR diversity and the identification of resistance genes involved in resistance to blast (*Magnaporthe oryzae*).

We have produced different genomic resources for this species which we have used to advance our understanding of the mechanisms involved in diplosporous apomixis as well as the identification of resistance genes for blast. Collectively, these resources contribute to the transition of *E. curvula* from orphan to well-characterized species.

PE0552: General Comparative

China National Genebank Database (CNCBdb)

Lijin You, China National GeneBank DataBase, shenzhen, China and **Lina Wang**, BGI, WUHAN, China

CNCBdb is a unified platform built for biological big data sharing and application services to the research community. Based on the big data and cloud computing technologies, it provides data services such as archive, analysis, knowledge search, management authorization, and visualization.

P00553: General Comparative

Genetic Contribution of Paleopolyploidy to Adaptive Evolution in Angiosperms

Yuannian Jiao, Institute of Botany, The Chinese Academy of Sciences, Beijing, China

Ancient whole-genome duplications (WGD or polyploidy) are prevalent in plants, and some WGDs occurred during the timing of severe global environmental changes. It has been suggested that WGDs may have contributed to plant adaptation. However, it still lacks of empirical evidence from genetic level to support the hypothesis. Here, we investigated the survivors of gene duplicates from multiple ancient WGD events on the major branches of angiosperm phylogeny, and aimed to explore genetic evidence supporting the significance of polyploidy. Duplicated genes co-retained from three waves of independent WGDs (~120 million years ago (Ma), ~66 Ma and <20 Ma) were investigated in 25 selected species. Gene families functioning in low temperature and darkness were commonly

retained gene duplicates after the eight independently occurred WGDs in many lineages around the Cretaceous–Paleocene (K-Pg) boundary, when the global cooling and darkness were the two main stresses. Moreover, the commonly retained duplicates could be key factors which may have contributed to the robustness of the critical stress related pathways. In addition, genome-wide transcription factors (TFs) functioning in stresses tend to retain duplicates after waves of WGDs, and the co-selected gene duplicates in many lineages may play critical roles during severe environmental stresses. Finally, our results shed new light on the significant contribution of paleopolyploidy to plant adaptation during global environmental changes in the evolutionary history of angiosperms.

PE0554: General Comparative

Comparative Transcriptome Analysis Reveals Conserved Genomic Characteristics of Drought Responsive Genes Across Multiple Plant Species

Jiyoung Lee, Lenwood S. Heath, Ruth Grene and Song Li, Virginia Tech, Blacksburg, VA

Drought stress is one of major abiotic stresses that negatively impacts agricultural production through slowing plant growth and reducing crop yield. In order to identify evolutionarily conserved drought responsive genes and their regulatory sequences, we have developed a unified computational pipeline to re-analyzed published RNA-seq data for drought responses across five plant species including *Arabidopsis*, soybean, poplar, rice and maize. Using conventional Venn Diagram analysis, we identified 120 common gene families induced by drought in these plant species. This is a surprisingly small number of gene families given that thousands of genes were affected by drought stresses in each species. To overcome the limitation of Venn Diagram analysis, we implemented a computational method to calculate branch length score (BLS) to quantify how drought responsive genes are conserved across species. BLS is used to visualize complex relations of subset of genes from multiple species with regard to their conserved expression patterns. Using BLS, we identified thousands of conserved gene families that are not identified by the previous Venn Diagram analysis. Taken together, the results of this study reveal novel common gene families with conserved drought responses, and conserved regulatory motifs across multiple species. The computational pipeline and the BLS approach developed in this study could be useful to annotate gene functions and understand underlying mechanisms of gene regulation in crop species under other abiotic and biotic stress conditions.

PO0555: General Comparative

Plant Telomerase RNA Genes: Conserved Motifs, Pol-III Transcription, and Scenario of Telomere Divergence

Petr Fajkus¹, Vratislav Peska¹, Eva Sýkorová¹, Michal Závodník² and Jiri Fajkus^{1,2}, (1)Institute of Biophysics, CAS, v.v.i., Brno, Czech Republic, (2)Masaryk University, CEITEC and Faculty of Science, Brno, Czech Republic

Telomerase RNA (TR) provides a template for telomere synthesis by telomerase reverse transcriptase (TERT). The sequence of telomere repeat is dictated by a template region (usually 1.5 telomere repeat long) of TR subunit, and is thus complementary to telomere sequence.

With recent characterisation of telomerase RNA genes in plants we described putative TRs from 75 representatives across land plant phylogeny. Four of them were experimentally validated as genuine telomerase RNAs in *Arabidopsis thaliana* (Brassicales), *Nicotiana sylvestris* (Solanales), *Allium cepa* and *Scilla peruviana* (both Asparagales). Despite of huge sequence variability between so far described 75 putative plant TRs, these sequences showed shared architecture of conserved motifs and promoter structure, typical for RNA polymerase III class III transcripts like U6, U3, 7SL, MRP RNAs involving Upstream Sequence Element (USE) and TATA box. Considering of extensive richness of Plantae kingdom, recently described TR sequences in 75 representatives (3 from Gymnosperms, 72 from Angiosperms) is just drop in the ocean.

Here we challenge Telomerase RNA genes in evolutionary ancestral plant clades outside seed plants (Spermatophyta), and clades where we previously failed with TR identification using Blast - these include clades Zingiberales and core Caryophyllales, involving important model and crop species.

We present a fresh comprehensive comparison of new RNA subunits from seed plants, ferns and lycophytes. Based on the upgraded comprehensive set of RNA subunits from seed plants, ferns and lycophytes we re-examine conserved TR motifs, which can reflect some functional significance in TR biogenesis.

In contrast to animals or yeasts, plant genomes frequently contain more TR-like genes. We hypothesize, that these TR paralogs can serve as a source material for telomere sequence evolution. Indeed, new screen for plant telomerase RNA subunits revealed several TRs harbouring template regions, which are not complementary to typical plant telomere motif (thus unable to synthesize typical plant telomere DNA). Two options can be considered – either some TR candidates identified *in silico* are not genuine telomerase RNAs or we find other species with unusual telomeric sequences. Fortunately, here we experimentally support the second option. Thus, we demonstrate the applicability of TR prediction for identification of evolutionary changes in plant telomeres.

Funding: Czech Science Foundation [17-09644S], ERDF [project SYMBIT, reg. no. CZ.02.1.01/0.0/0.0/15 003/0000477]

PE0556: General Comparative

Understanding the Role of TIP3 Aquaporins in Plant Seed Development

Praveen Khatri, NATIONAL AGRI-FOOD BIOTECHNOLOGY INSTITUTE, MOHALI, INDIA, MOHALI, India

Aquaporins (AQPs), channel forming proteins, are of vital importance in the cellular transport system of almost all organisms. In plants, AQPs are known to transport water and many other small solutes like urea, H₂O₂, CO₂, and metals like aluminium and metalloids including boron, silicon, arsenate, and germanium. Regulation of water transport by AQPs is highly studied aspect concerning its effect on abiotic and biotic stress in plants. Several studies have highlighted AQPs role in water and nutrient uptake in root and subsequent transport to the aerial part of the plant. However, insufficient information is known about AQPs transport system in seed development. Earlier reports by our group have shown expression dynamics of AQP genes during different stages of seed development. We have identified Tonoplast Intrinsic Proteins (TIPs) a subfamily of aquaporins as a candidate involved in the seed desiccation process. Interestingly, all the AQPs except TIP3s have been found to be downregulated during seed development in canola, soybean, rice, flax, and Arabidopsis. In the present study, to confirm the role of TIPs more particularly TIP3s, we are analyzing seed development in Arabidopsis TIP3 mutant. A TIP3 (GmTIP3-1) cloned from soybean is being studied in Yeast Complementation assay to confirm solute specificity and transport kinetics. In addition, we will perform transcriptome profiling and gene co-expression analysis of the mutants. Similarly, knockout mutants for soybean TIP3s are being generated through the genome-editing approach. Understanding of seed desiccation will be helpful to enhance seed viability, storage ability, as well as nutrient content more particularly seed oil.

PO0557: General Comparative

Functional Divergence of Duplicate Genes through Exon-Intron Structural Changes

Xuehao Fu, Hongyan Shan, Peipei Wang, Haifeng Zhao, Meimei Wang, Hongzhi Kong and **Guixia Xu**, Institute of Botany, Chinese Academy of Sciences, Beijing, China

Gene duplication provides raw genetic materials for the origin of new genes and new characters. Despite extensive studies, it remains unclear how duplicate genes diverge in function. Previous studies have shown that, at least in plants, changes in exon-intron organization (structural divergence) could be an important mechanism for the divergence of duplicate genes. However, because the species and gene pairs sampled were limited, it is yet unclear whether this phenomenon is widespread. In this study, by conducting genome-wide analyses on closely-related duplicate genes from representative species of plants (*Arabidopsis thaliana*), animals (*Drosophila melanogaster*), fungi (*Saccharomyces cerevisiae*), and protists (*Paramecium tetraurelia*), we found that structural divergence occurred prevalently in every examined species but with different proportions, ranging from 71% in *P. tetraurelia* to 91.9% in *A. thaliana*. Three mechanisms, including exon/intron gain/loss, exonization/pseudoexonization, and intra-exonic insertion/deletion, are detected to make unequal contributions to structural divergence. The occurrence probability and frequency of exonization/pseudoexonization and intra-exonic insertion/deletion are positively

correlated with evolutionary time, but negatively correlated with functional constraint. We also found that the Pearson correlation coefficient (PCC) of expression is lower between structurally diverged than undiverged genes, suggesting that structural divergence may be coupled with expression divergence. More importantly, in comparison of orthologs and duplicate genes with similar evolutionary time, duplicate genes tend to accumulate structural changes more easily and more frequently. These findings altogether show that the modes of structural divergence of duplicate genes are generally consistent in different eukaryotic species, implying that structural divergence is an important contributor to the evolution of duplicate genes.

PE0558: General Comparative

Evolutionary Changes of the Opsin Genes in 250 Vertebrates and Their Implications for Environmental Adaptation

Jinn-Jy Lin¹, Feng-Yu Wang², Tzi-Yuan Wang¹ and Wen-Hsiung Li¹, (1)Academia Sinica, Taipei, Taiwan, (2)Taiwan Ocean Research Institute, National Applied Research Laboratories, kaohsiung, Taiwan

Vertebrates use opsin to sense light in terrestrial and aquatic environments which differ greatly in the availability of light of different wavelengths. However, how their vision has adapted to such diverse living environments in the course of vertebrate evolution is still underexplored. To address this issue, we carefully identified the opsin genes in the genomes of 250 vertebrates and studied the opsin gene trees, genes adjacent to opsin genes (syntenies) and amino acids at key residues (tuning sites) in these species. We then inferred the evolutionary changes in opsin gene copy number, syntenies and tuning sites along each lineage. We found that Actinopterygii species (ray-finned fishes) have massive duplications and losses of opsin genes, while Sauropsida species (turtles, lizards, snakes and crocodilians) have no duplication but only differential gene loss events. The syntenies are highly conserved in mammals but some rearrangements have occurred in Sauropsida and massive rearrangements have occurred in Actinopterygii. Finally, we discussed the implications of the evolutionary changes in opsin gene copy number and in amino acids at tuning sites for the visual adaptation of vertebrates to their diverse living environments.

PO0559: General Comparative

The Genomic Impact of Hybridization in Old World *Myotis* Bats; Implications for Phylogenomics.

Nicole Foley and William J. Murphy, Texas A&M University, College Station, TX

Emerging phylogenomic studies are beginning to reveal the confounding role that genome architecture and introgression play in the most difficult of phylogenomic problems, resolving relationships among hybridizing species. The dominant tree arising from simultaneous analyses of entire genomes is often considered the species tree. Recent studies of hybridizing species have challenged this assumption, instead showing that the species tree is often not the dominant signal, is localized in low recombining regions of the genome and areas enriched for species barriers (e.g. X chromosome). *Myotis* bats represent one of the most successful mammalian radiations, notable for high species numbers (>100), almost worldwide distribution, extreme morphological convergence, exceptional longevity and status as viral reservoirs. A resolved phylogeny is key to understanding the evolution of these bats' unique traits, an objective hampered by ongoing hybridization. To address this, we generated novel whole genomes from 60 museum samples collected throughout the Old World, representing at least 40 *Myotis* species. All data were mapped to a *Myotis myotis* chromosome level assembly. A maximum likelihood sliding window approach was used to characterize variation in phylogenomic signal across the genome and to determine if the distribution of phylogenomic signal is linked to historical recombination rate. These data will characterize introgressed regions of the genome and overcome barriers to phylogenomic resolution among these bats. Results from this study provide an evolutionary context to fully understand the unusual adaptations that have arisen through a history of introgression in *Myotis* bats.

PE0560: General Comparative

Divergence of *Aspergillus flavus* Morphotypes and Closely Related *Aspergillus oryzae* Groups Based on Genome-Wide Nucleotide Variations

Perng-Kuang Chang, SRRC/ARS/USDA, New Orleans, LA, Thomas D. Chang, Software Engineer, New Orleans, LA and Kazutaka Katoh, Osaka University, Suita, Japan

Genome sequences of 12 *Aspergillus flavus* and 11 *Aspergillus oryzae* strains were retrieved from the National Center for Biotechnology Information database. The sequences were analyzed by Mauve, a multiple-genome alignment program. Genome-wide nucleotide variations (total SNPs) were extracted from strain comparison of the same species (*A. flavus* vs *A. flavus* and *A. oryzae* vs *A. oryzae*) or between the two species (*A. flavus* vs *A. oryzae*). Results showed that averaged total SNPs among *A. flavus* strains of the same morphotype (L or S) were consistently lower than those between different morphotype strains (L and S). Phylogenetic analysis with concatenated total SNP sequences by the distance-matrix UPGMA method of the online program MAFFT (version 7) inferred that the L-morphotype and *A. oryzae* diverged from the S-morphotype 5.2 mya (million years ago). The L-morphotype and *A. oryzae* groups 1 and 2 were separated into two clades. *A. flavus* NRRL21882, the active agent of the biocontrol product Afla-Guard® GR, also diverged from the S-morphotype at about the same time (5.1 mya). Another EPA-registered biocontrol strain, *A. flavus* AF36, diverged from toxigenic L-morphotype strains 2.6 to 3.0 mya. The close genetic relatedness of *A. flavus* to *A. oryzae* was confirmed, and the evolutionary origins of currently used atoxigenic *A. flavus* biocontrol strains were revealed.

PO0561: General Comparative

Submission, Archival and Visualisation of Single-Cell Sequencing Data

Nancy George, European Bioinformatics Institute (EMBL-EBI), Hinxton, United Kingdom and Irene Papatheodorou, European Bioinformatics Institute, Hinxton, United Kingdom

Single Cell Expression Atlas (SCEA) (www.ebi.ac.uk/gxa/sc) is the latest component of the Expression Atlas. The Expression Atlas knowledgebases form part of a comprehensive suite of functional genomics resources developed at the EBI from submission to visualisation. SCEA ingests and systematically reanalyses and visualises publicly available single cell RNA-sequencing (scRNA-seq) datasets from functional genomics archives such as ArrayExpress; NCBI's GEO and ENA. As of its latest release (December 2019), SCEA contains 132 datasets, across 12 species and over 1.3 million cells.

For inclusion into the Single Cell Expression Atlas all datasets are curated to a high standard complying with minimum metadata requirements (described in detail here: <https://arxiv.org/abs/1910.14623>). These standards are incorporated into the single cell submission template in our web submission tool Annotare. Upon acceptance, datasets are given a stable, citable accession, reviewed by curators and uploaded to ArrayExpress – a functional genomics data archive. ArrayExpress (www.ebi.ac.uk/arrayexpress) then displays each dataset as a self-contained entity providing the experiment information; sample metadata and links to raw and processed data under a single accession.

From other resources, such as NCBI's GEO or EGA, when a suitable dataset is discovered, raw data is retrieved from the corresponding archive or data submitter. Sample metadata is curated manually and where possible annotated to ontology terms for easier data search and retrieval. All raw data is processed through standardised analysis pipelines depending on the scRNA-seq technology used. All analysis workflows are available through our GitHub repositories: <https://github.com/ebi-gene-expression-group/> whilst tools to run these are made available via Galaxy here: <https://humancellatlas.usegalaxy.eu>.

Once reanalysed, the scRNA-seq analysis results are made available to the wider scientific community through the visualisations in the SCEA interface. Through this platform users can search for genes across datasets and filter the results for particular cell types or tissues. SCEA can also be used to identify in what conditions and populations a gene can act as a marker gene, i.e. define a specific cell population. For each experiment cell populations are displayed via a t-SNE plot. Cells can also be coloured with the underlying metadata. Gene expression at the single cell level can be explored in the neighbouring plot. The top 5 marker genes per cluster are also displayed and all analysis data and accompanying metadata are available for download.

PE0562: Brassicas, Arabidopsis, and related

Upgrade the Search Function of the Plant Bioresource Database, Exp-Plant Catalog, in RIKEN BRC

Satoshi Iuchi and Masatosmo Kobayashi, RIKEN Bioresource Research Center, Tsukuba, Japan

The RIKEN BioResource Research Center (RIKEN BRC) Experimental Plant Division participates in the National BioResource Project (NBRP) and collects, maintains and provides plant resources, such as wild-type and mutant seeds of *Arabidopsis thaliana* as well as full-length cDNA clones and plant cultured cell lines of various plant species.

We have developed a web catalog (Exp-Plant Catalog) that provides the information of these resources. In the Exp-Plant Catalog, the wild-type lines of *A. thaliana* are accompanied with not only the basic information such as location of origin but also the phenotypic information including photo images. For *A. thaliana* transposon-tagged mutants, the insertion site of transposable element is displayed as a diagram on the simplified genome browser. The genome browser is also available for the cDNA clones of *A. thaliana*. Users are able to obtain the information of the bioresource of their interests by searching entire Catalog or part of Catalog by limited by the type and species.

In order to upgrade the search function of the Exp-Plant Catalog, we intend to introduce multiple word search. As this change includes the full-text search of gene annotation, it will reduce the performance of search. In order to recover the performance of search, we will process the text of gene annotation. Currently we are trying to improve the user interface in the web pages which will be provided from the Exp-Plant Catalog (<https://plant.rtc.riken.jp/resource>).

PO0563: Brassicas, Arabidopsis, and related

Experimental Hybrid Speciation in *Brassica*

Elvis Tembang Katche, Department of Plant Breeding, Justus-Liebig University Giessen, Giessen, Germany, Roman Gaebelien, Justus Liebig University Giessen, Giessen, Germany, Jacqueline Batley, University of Western Australia, Perth, WA, Australia and Annaliese S Mason, Plant Breeding Department, IFZ Research Centre for Biosystems, Land Use and Nutrition, Justus Liebig University Giessen, Giessen, Germany

Interspecific hybridisation is an important path to generating evolutionary novelty, and the genus *Brassica* is a major agricultural genus with a history of interspecific hybridisation. The allotetraploids *B. juncea*, *Brassica carinata*, and *Brassica napus* were formed by pairwise hybridisation between diploid progenitors *Brassica rapa*, *Brassica oleracea* and *Brassica nigra*. Although crossing between these allotetraploid species is possible and has been carried out in several studies either to study chromosome behavior or to transfer useful traits between species, attempts to generate novel, stable and fertile synthetic hybrids through this method have not been reported. We generated interspecific hybrids (AABC = $F_1 = 37$), by crossing *B. juncea* ($2n = AAB = 36$) \times *B. napus* ($2n = AAC = 38$), (CCAB = $F_1 = 36$) by crossing *B. napus* ($2n = AAC = 38$) \times *B. carinata* ($2n = BBCC = 34$) and (BBAC = $F_1 = 35$) by crossing *B. juncea* ($2n = AAB = 36$) \times *B. carinata* ($2n = BBCC = 34$) and self-pollinated these hybrids for generations by selecting for fertility. CCAB and AABC hybrids became infertile in the early generations (S_1 and S_2 respectively) while BBAC increased in fertility across generations (up to S_6). In the absence of homologous pairing partners, the A and C genomes paired, restructured and stabilized to form viable and fertile offspring. This pathway can be useful for generating evolutionary novelty which can be transferred to other *Brassica* species and also to produce new useful crop types.

PE0564: Brassicas, Arabidopsis, and related

Identification and Characterization of Superoxide Dismutase (SOD) Gene Family Associated with Abiotic-Stress Responsive in Brassica Species (*B. juncea*, *B. rapa* and *B. nigra*)

Deepika Verma, Neha Lakhanpal and Kashmir Singh, Panjab University, Chandigarh, India

Background: Abiotic stresses cause major productivity loss in the rapeseed-mustard crops (*Brassica*). Efforts in the past have been made to identify such gene families that will help provide resistance against these stresses. Cellular metabolism produces reactive oxygen species (ROS) that lead to cell death. To deal with such ROS species,

organisms have developed antioxidative defense mechanisms that comprise of enzymes such as Superoxide dismutase(SOD), Catalase (CAT), glutathione peroxidase (GPX), ascorbate peroxidase (APX), and glutathione reductase (GR). SODs are metalloenzymes that constitute the first line of defense against ROS and their classification is done according to their metal species which catalyzes the dismutation reaction $O_2^{\cdot -} + 2H^+ \rightarrow O_2 + H_2O_2$. The present study reports genome-wide identification of abiotic stress responsive *SOD* gene family in *B. juncea* and *B. rapa*.

Results: We identified total 29,18 and 16 *SOD* genes in *B. juncea*, *B. rapa* and *B. nigra* respectively. These genes were mapped on the chromosomes manually. Phylogenetic classification of all these genes was based on their domain composition, which was also supported by sub-cellular locations predictions. For the annotation prediction of *SODs* Gene Ontology (GO) mapping was done via BLAST2GO and the result was supported by the *cis*-regulatory elements predicted in the promoter region of *SOD* genes. FPKM analysis was done using SRA data available for drought, heat and salt stress. A total of 10 and 14 abiotic stress responsive *SOD* genes were successfully predicted in *B. juncea* and *B. rapa* respectively. Validation of candidate responsive genes under drought and heat stress was done through quantitative Real Time PCR.

Conclusion: To aid the improvement of plant tolerance against various abiotic-stresses, gene families were identified. Our study will help in providing significant information of various ROS scavenging gene families linked to abiotic stresses which would act as a potential resource for the improvement of stress resistance in *Brassica* crops.

PO0565: Brassicas, Arabidopsis, and related

Enhancement of Experiments in Artificial Selection: QTL Analysis of Loci That Condition Expression of Variable Traits in Rapid-Cycling *Brassica Rapa*

Scott Woody, UW-Madison, Madison, WI

Experiments in artificial selection are a staple of the biology laboratory curriculum, but students are only rarely provided insight into the nature of the loci and associated allelic variants that might account for the response to selection they imposed on parental populations. To overcome that deficiency, we conducted QTL analysis of phenotypically variable traits in *Brassica rapa* that are readily quantified and so suitable as targets for selection. Specifically, we used two *B. rapa* RIL populations (R500 x IMB211 and R500 x FPsc) to identify loci where allelic variation conditions expression of seed coat color (*SCC*), anthocyanin pigments in hypocotyls (*AN*), trichome abundance on leaf margins (*TC*), flowering time (*DTF*) and height to first flower (*HtFF*). Our consideration of QTL peak data was aided considerably by the availability of *de novo* sequence assemblies of both FPsc and the common maternal parent R500. As expected, QTL peaks were evident on in both RIL populations, consistent with prior evidence showing that insertion of a *HELITRON* transposable element in a *B. rapa* orthologue of Arabidopsis *TRANSPARENT TESTA8* (*TT8*) largely accounts for the bright yellow seeds of R500 and the brown seeds of IMB211 and FPsc parental varieties. Similarly, both RIL populations yielded *AN* QTL peaks near an ortholog of Arabidopsis *PRODUCTION OF ANTHOCYANIN PIGMENTS2* (*PAP2*). However, the genetic architecture underlying expression of *TC*, *DTF*, and *HtFF* traits were surprisingly divergent. The results of all QTL analyses will be described, as well as the potential of these new data to enhance student learning outcomes.

PE0566: Brassicas, Arabidopsis, and related

Genome-Wide Identification of Resistance Gene Analogs in the Brassicaceae

Soodeh Tirnaz¹, Philipp E. Bayer¹, Fabian C. Inturrisi², Fangning Zhang¹, Hua Yang³, Aria Dolatabadian², Ting Xiang Neik², Anita Severn-Ellis¹, Aneeta Pradhan², David Edwards¹ and **Jacqueline Batley¹**, (1)University of Western Australia, Perth, WA, Australia, (2)University of Western Australia, Perth, Australia, (3)University of Queensland, Brisbane, UQ, Australia

The Brassicaceae consists of a wide range of species, including agronomically and economically important Brassica crop species, however their production is affected by diseases causing significant yield losses annually worldwide. A major way to improve resistance is through identification, selection and implementation of resistance gene analogs (RGAs) in breeding. Nucleotide-binding site-leucine-rich repeat (NLR), receptor like kinases (RLK) and receptor like proteins (RLP) genes are the main types of RGAs. They contain conserved domains and motifs and play

specific roles in resistance to pathogens. Here, all RGA classes were identified among all currently available genome annotations from the Brassicaceae, as well as among different genome assemblies of the same. The results show the number of RGAs varies within and between species. In total 34,065 RGAs were identified. To evaluate the impact of annotation methods on RGA predictions NLRs were also identified independent of annotation, producing similar results. This study provides a valuable resource for the identification and characterization of RGAs in the Brassicaceae to assist Brassica breeders toward resistance improvement.

PO0567: Brassicas, Arabidopsis, and related

Systems Genetics Analysis of Waterlogging Tolerance Mechanisms in Brassica

Jheng-Yang Ou¹, Po-Xing Zheng¹, Ta-Yu Yang¹, Wang Lin¹, Chen-Yu Lin² and **Yao-Cheng Lin¹**,
(1)Biotechnology Center in Southern Taiwan & Agricultural Biotechnology Research Center, Academia Sinica, Tainan, Taiwan, (2)Department of Vegetable Crops, Fengshan Tropical Horticultural Experiment Branch, Taiwan Agricultural Research Institute, Kaohsiung City, Taiwan

Due to their sessile nature, plants are unable to escape from stressful environment that prevents the plants from attaining their full genetic potential for growth and reproduction. Environmental variables, especially those affecting temperature and water availability, are major determinants of plant growth and reproduction. Understanding the genetic basis of phenotypic variations is the central aim in many biological research questions. Modern bioinformatics and genomics take advantages of high throughput sequencing that genome sequences and changes of gene expression values responding to stimuli of non-model organisms can be measured cost effectively. In this study, we integrate genetics, bioinformatics and phenomics to help us understanding genes that are involved in waterlogging stress responses. Using a biparental mapping population including two parental lines, F1 hybrid and the F2 population, plant phenotypes including number of leaves, plant height and projected leaf area were routinely recorded during the growth period. Applying a four parameter Gompertz growth model, we were able to predict the plant height inflation point where the plants reach its maximum height and likely shift from vegetative growth into reproductive growth. Forty days after transplanting, plants were treated with 48 hours of waterlogging and the damaged leaf area were scored and gene expression values were measured by RNAseq. Based on the bulk segregant analysis by integrating gene expression values and phenotypes, we have identified genes that are involved in hypoxia responses, photosynthesis and hormone signaling might play important role in waterlogging tolerance.

PE0568: Brassicas, Arabidopsis, and related

The WRKY Gene Family in Ramie (*Boehmeria nivea* L. Gaud): 26 BnWRKY Gene's Expression Analyses and Overexpression of BnWRKY23 Increased Drought Stress Tolerance in *Arabidopsis thaliana*

Yaning Bao, Huazhong Agricultural University, Wuhan, China

Ramie (*Boehmeria nivea* L. Gaud), which is also known as “China grass”, originated in China and is a perennial herbaceous plant of the Urticaceae family. Ramie fiber made from stem bast is a good textile material and the leaves and roots are used in Chinese medicine. The *WRKY* family is one of the largest transcription factor families in plants, and its members play important roles in plant growth and development and multiple stress responses. In this study, 26 *WRKY* genes (named *BnWRKY1*-*BnWRKY26*) were identified in ramie and classified into three main groups by analysis of their gene structure and conserved motifs. Expression profiles of *BnWRKY* genes at three stages of ramie growth and in response to abiotic stresses (salt and drought) were analyzed. The results revealed that the *BnWRKY* genes exhibited differential expression at the three stages and were induced by salt and drought stresses. In addition, overexpression of *BnWRKY17* and *BnWRKY23* in *Arabidopsis* resulted in an increased tolerance to drought and salt, and *BnWRKY23*-overexpressing lines exhibited decreased sensitivity to ABA during early seedling growth. Moreover, the *BnWRKY23*-overexpressing lines exhibited increased tolerance to drought stress because of decreased stomatal size and increased cuticular wax deposition on their stems. In summary, these results reveal that heterologous expression of *BnWRKY23* in *Arabidopsis* increased tolerance to salinity and drought stress. This study will help understand the potential molecular response mechanism for drought stress and will provide assistance to the breeding of drought-resistant in ramie.

PO0569: Brassicas, Arabidopsis, and related

A Chromosomal-Scale Genome Assembly of *Arabidopsis halleri* ssp. *halleri*

Bjoern Pietzenuk, Ruhr-University Bochum, Bochum, NRW, Germany

Arabidopsis halleri is a self-incompatible, stoloniferous, perennial outcrossing diploid ($2n = 16$) plant with an approximate genome size of 256 Mbp. The *A. halleri* lineage is closely related to the classical genetic model *A. thaliana* and the *A. lyrata*. *A. halleri* diverged from *A. thaliana* ~5 Mya followed by its divergence from *A. lyrata* ~3 Mya after which the trait of heavy metal accumulation evolved¹.

A metal hyperaccumulator plant can accumulate and tolerate metal concentrations more than an order of magnitude above the critical toxicity thresholds of ordinary plants. Thus, metal hyperaccumulation requires extraordinarily efficient and specific mechanisms of metal extraction from the soil, root-to-shoot metal partitioning, and internal metal detoxification. To date, metal hyperaccumulation has been identified in more than 500 plant taxa, corresponding to about 0.5% of all vascular plants.

As a hyperaccumulator of Zn and Cd, *A. halleri* tolerates up to 76-fold higher zinc and 8-fold higher cadmium concentrations in soil than *A. thaliana*^{2,4}. Metal hyperaccumulation and associated hypertolerance in plants bear great promise for the development of phytoremediation and phytomining technologies, in addition to serving as extreme model traits in molecular physiology, evolution and ecology. *Arabidopsis halleri* is distributed across Eastern and Central Europe. In a field survey, *A. halleri* genotypes were collected at 165 collection sites of which only 59 were metal contaminated soils. To study the natural variation of metal accumulation and tolerance between *A. halleri* accessions, suitable reference genome assembly is required. Previously, published genome assemblies were of insufficient quality and contiguity. We will present the present *A. halleri* subsp. *halleri* genome assembly from Langelshiem, Germany (Lan3.1), covering 196 Mbp at the pseudochromosomal scaffold level.

1) Clauss MJ, Koch MA. 2006. Poorly known relatives of *Arabidopsis thaliana*. *Trends Plant Sci.* 11:449–59

2) Willems G, Draßger DB, Courbot M, Godec C, Verbruggen N, Saumitou-Laprade P. 2007. The genetic basis of zinc tolerance in the metallophyte *Arabidopsis halleri* ssp. *halleri* (Brassicaceae): an analysis of quantitative trait loci. *Genetics* 176:659–74

3) Bert V, Meerts P, Saumitou-Laprade P, Salis P, Gruber W, Verbruggen N. 2003. Genetic basis of Cd tolerance and hyperaccumulation in *Arabidopsis halleri*. *Plant Soil* 249:9–18

PE0570: Brassicas, Arabidopsis, and related

Genetic Components for Acidity Tolerance in Arabidopsis

Menuka Bhandari, Arjun O. Kshetry, Suhas Shinde, Carlos Ortiz, Thangasamy Saminathan, Padma Nimmakayala and Umesh K. Reddy, Department of Biology, West Virginia State University, Institute, WV

Acidic soil constitutes around forty percent of the global arable land. This phenomenon can lead to toxicity and deficiency of some essential elements, which results in low productivity. In order to elucidate molecular mechanisms and identify genes for acidity tolerance, we conducted a GWAS involving natural variation of 244 *Arabidopsis thaliana* ecotypes to identify 725 significant SNP associations that resolved into 110 candidate genes. We prioritized 38 genes for reverse genetic analysis using publicly available knockout mutants to functionally validate their role in acidity tolerance. The loss-of-function mutant analysis revealed two important genes, i.e., RB1 superfamily protein and AB1 superfamily protein. Further validation of these two genes by over-expression and complementation is in progress.

PO0571: Brassicas, Arabidopsis, and related

Exploration into Natural Variation for Arsenic Effects in Arabidopsis

Yadira Peña-García¹, Cristina Delgado-Chavez¹, Carlos Ortiz¹, Nagamani Balagurusamy², Suhas Shinde¹, Thangasamy Saminathan¹, Padma Nimmakayala¹ and Umesh K. Reddy¹, (1)Department of

Biology, West Virginia State University, Institute, WV, (2)Faculty of Biological Sciences, Autonomous University of Coahuila, Torreon, CU, Mexico

A better understanding of Arsenic tolerance mechanisms might lead to application of biotechnological solutions for developing As-resistance crops. Previous studies have identified several As responsive genes involved in its metabolism, translocation, sequestration, etc., however, many of the main key players remain elusive. In order to identify additional loci involved in As stress, we used genome-wide association analysis to study the natural variation of seedling root traits in a panel of 166 Arabidopsis accessions exposed to As (V) and control. GWAS resulted in a total of 770 significantly associated SNPs, which corresponded to 277 genes. Further, a subset of 68 candidate genes were prioritized and used for functional characterization through reverse genetic approach. Compared with wild-type, one of the T-DNA insertion mutants (*sae24*) displayed As susceptibility with a relative root length of 51% reduction. Currently, we are investigating the role of SAE24 in the As tolerance by over-expression and creating double mutants with known genes.

PE0572: Brassicas, Arabidopsis, and related

The Role of Light and Phytoglobins in Arabidopsis Somatic Embryogenesis

Samuel A. Day¹, Claudio Stasolla¹, Mohammed M Mira¹ and Robert D Hill², (1)University of Manitoba, Winnipeg, MB, Canada, (2)University of Manitoba, Winnipeg, MB, Canada

Somatic embryogenesis (SE), the process of generating plant embryos from somatic cells, is a useful tool for studying factors which impact early embryo development in plants, including Arabidopsis. Experimentally, SE is broadly divided into two stages. During the induction phase, embryogenic tissue is grown from immature, green zygotic embryos with the help of auxin (2,4-D) in the growth media. This is followed by the maturation phase, characterised by the development of embryos from the embryogenic tissue in the absence of hormones. Embryogenic tissue proliferation during SE is controlled by jasmonic acid via regulation of auxin biosynthesis. Phytoglobins (Pgbs) are heme-group containing proteins which scavenge nitric oxide (NO). Of the phytoglobins found in Arabidopsis: Pgb1 and Pgb2, suppression of the former inhibits SE while suppression of Pgb2 promotes the formation of the embryogenic tissue by inducing JA and auxin synthesis via NO modulation. Jasmonic acid synthesis is closely linked to light, and early results show a positive effect of light on embryo yield and induction of auxin biosynthetic genes. Additionally, results suggest an interaction between light and phytochrome expression during the induction phase, while the critical step of embryogenic tissue formation is taking place. White light induces Pgb1 expression. Darkness induces Pgb2 expression during the induction phase while reducing Pgb1 expression seen under white light exposure. Moving forward, our objective is to elucidate the link between light and phytochrome expression during somatic embryogenesis.

PO0573: Brassicas, Arabidopsis, and related

Lack of the Q-Type C2H2 Zinc Finger Transcription Factor Zat18 Alters the Response to Infestation in Arabidopsis

Susan D. Lawrence and Nicole G. Novak, USDA-ARS, Beltsville, MD

Q-type C2H2 zinc finger proteins (ZFPs) play a role in the plant response to stress. Overexpression of individual ZFPs resulted in enhanced tolerance to abiotic stresses such as drought or salt. Overexpression of ZFP, Zat18 enhanced tolerance to drought, while loss of Zat18, produced a plant more vulnerable to water loss. The effect of Zat18 loss on gene expression has not been characterized. Overexpression of an ortholog of Zat18 in potato, StZFP2 enhances tolerance to infestation by *Manduca sexta*. This work focuses on two Zat18 mutant lines, a knockout and another with a residual amount (knockdown) of Zat18. The two lines show differences in response to infestation. Typically, upon infestation by a chewing insect, genes in the jasmonic acid (JA) pathway are induced. While infestation does induce JA pathway genes, Zat18 mutant lines down regulate ABA signaling genes and up regulate ethylene signaling genes ERF1 and ERF2 as well as SA pathway gene NPR3. Another interesting discovery is that the knockdown Zat18 mutant line with a residual amount of Zat18 expression causes far more changes in the infestation response compared with Col-0 or the Zat18 knockout line.

PE0574: Brassicas, Arabidopsis, and related

Involvement of the Atypical bHLH Transcription Factor Paclobutrazol Resistances in ABA and Salt Responses in Arabidopsis

Kaijie Zheng, Northeast Institute of Geography and Agroecology, CAS, Changchun, China

Absciscic acid (ABA) plays an important role in regulating plant response to abiotic stresses. Several bHLH transcription factors have been reported to regulate ABA signaling via different ways in Arabidopsis. Paclobutrazol Resistances (PREs) are non-DNA binding bHLH transcription factors. They have been reported to involve in the regulation of plant response to several different plant hormones and environmental stimuli including gibberellin, brassinosteroid, auxin and light. Here, we show that PREs are involved in the regulation of ABA and salt responses in Arabidopsis. Quantitative RT-PCR results showed that the expression levels of *PRE2* and *PRE6* as well as several other *PRE* genes were reduced in response to ABA treatment, but increased to salt treatment. Seed germination assays indicated that ABA sensitivity is reduced in the *pre2*, *pre6* and *pre2 pre6* mutants, but increased in transgenic plants overexpressing *PRE2* and *PRE6*, respectively. We also found the *35S:PRE2* and *35S:PRE6* transgenic plants showed enhanced tolerance to salt, whereas little, if any changes were observed in the single and double mutants. Taken together, our results suggest that at least some of the PRE genes are ABA responsive genes. And the slight changes in seed germination and salt treatment experiment in double mutants suggests that PREs may function redundantly to regulate ABA and salt responses in Arabidopsis.

Keywords: Paclobutrazol resistances, bHLH transcription factor, Absciscic acid, Abiotic stress, Salt tolerance

PO0575: Brassicas, Arabidopsis, and related

Transcriptome Profiling of Seed Developmental Stages in *Arabidopsis thaliana*

Minsu Park, Seoul National University, Seoul, South Korea, Chanseok Shin, Seoul National University, Seoul, Korea, Republic of (South) and Authors' Group

Plant seeds are an important genetic mediator for the trait of progeny. Seed development is related with germination which is regulated by seed dormancy. Seed dormancy is affected and regulated by a large number of gene regulation and environmental condition. The antagonistic interaction between absciscic acid and gibberellin plant hormones plays an important role in the regulation of seed dormancy and germination. Recent studies of epigenetics have shown that seed dormancy and germination-related gene expression is regulated by DNA methylation and histone acetylation. However, the molecular mechanism underlying the process during seed development has not been clearly understood. In this study, we analyzed transcriptome and small RNAome of three different dormancy levels of seed development stages in Arabidopsis (*Arabidopsis thaliana*) ecotype Col-0. We investigated transcriptome-wide responses of small RNAs, long non-coding RNAs (lncRNA), and protein-coding genes during breaking seed dormancy. We found that differential expression of well-known marker genes associated with the processes of dormancy and germination in Col-0 such as *Dog1*, *ABI5*, *Ga3ox1* and *Ga3ox2*. To verify these data, we analyzed expression level of these marker genes in each different dormancy stage by quantitative reverse transcription PCR. Also, we identified approximately 1,500 lncRNAs and some of them located near genes which are related to dormancy. This study provides a foundation for understanding dynamics of transcriptome during breaking seed dormancy.

This work was supported by the Next-Generation BioGreen 21 Program (No. PJ01332501), Rural Development Administration, Republic of Korea.

PE0576: Brassicas, Arabidopsis, and related

Three CNGC Family Members, CNGC5, CNGC6 and CNGC9, Are Required for Constitutive Growth of Arabidopsis Root Hairs As Ca²⁺-Permeable Channels

Yan-Qiu Tan, Shanghai Institute of Plant Physiology & Ecology, SIBS, CAS, Shanghai, China, Yang Yang, Shanghai Institute of Plant Physiology & Ecology, CAS, Shanghai, China, Shanghai, China and **Yong-Fei Wang**, Institute of Plant Physiology and Ecology, CAS, China, Shanghai, China

Root hairs are tubular-shaped structures resulted from the outgrowth of root epidermal cells, greatly increase the overall surface area of roots, and are essential for the acquisition of diverse ion nutrients and water as well as

interaction between plants and microbes in rhizosphere. It has been well established that a cytosolic Ca^{2+} gradient and its oscillation at root hair apex are required for the polar growth and orientation of root hairs. Plasma membrane Ca^{2+} channels at root hair tips have been hypothesized as the key components for the establishment and dynamic regulation of the Ca^{2+} gradient via mediating and regulating the influx of external Ca^{2+} , and are consequently essential for root hair growth. However, the genetic identities of the Ca^{2+} channels in root hair tips essential for constitutive root hair growth have remained elusive for decades. Here, we report the identification and characterization of three cyclic nucleotide-gated channel (CNGC) family members, CNGC5, CNGC6, and CNGC9, as Ca^{2+} channels essential for constitutive root hair growth in *Arabidopsis*. We found that the *cngc5-1 cngc6-2 cngc9-1* triple mutant (designated *shrhl*) showed significantly shorter and branching root hair phenotypes as compared with the wild type. The defective root hair growth phenotype of *shrhl* could be rescued by either the expression of CNGC5, CNGC6, or CNGC9 single gene or by the supply of high external Ca^{2+} , but could not be rescued by external K^{+} supply. Cytosolic Ca^{2+} imaging and patch-clamp data in HEK293T cells showed that these three CNGCs all function as Ca^{2+} -permeable channels. Cytosolic Ca^{2+} imaging in growing root hairs further showed that the Ca^{2+} gradients and their oscillation in root hair tips were dramatically attenuated in *shrhl* compared with those in the wild type. Phenotypic analysis revealed that these three CNGCs are Ca^{2+} channels essential for constitutive root hair growth, with different roles in root hairs from the conditional player CNGC14. Moreover, we found that these three CNGCs are involved in auxin signaling in root hairs. Taken together, our study identified CNGC5, CNGC6, and CNGC9 as three key Ca^{2+} channels essential for constitutive RH growth and auxin signaling in *Arabidopsis*. This work was recently published in a new launched journal called *Plant Communications* (Tan et al., 2019), and this paper was recommended by Faculty 1000 as a “Very Good” paper with two stars.

Reference

Tan, Y.-Q., Yang, Y., Zhang, A., Fei, C.-F., Gu, L.-L., Shu-Jing, S., Xu, W., Wang, L., Liu, H., and Wang, Y.-F. (2019). Three CNGC family members, CNGC5, CNGC6, AND CNGC9, are required for constitutive growth of *Arabidopsis* root hairs as Ca^{2+} -permeable channels. *Plant Communications*, Doi: 10.1016/j.xplc.2019.100001.

PO0577: Brassicas, *Arabidopsis*, and related

Phenotypic Characterization of Mitochondrial Uncoupling Protein (UCP) T-DNA Mutants of *Arabidopsis thaliana* Reveals Severe Impact on Plant Fertility

Ivan G. Maia¹, Mariana de Lara Campos Arcuri², Rômulo P. M. Lima², Alessandra Vasconcellos Nunes Laitz³ and Pedro Barreto⁴, (1)Biosciences Institute of Botucatu - Department of Genetics - UNESP, Botucatu, SP, Brazil, (2)Biosciences Institute of Botucatu - Department of Genetics - UNESP, Botucatu, Brazil, (3)Instituto Federal de Educação, Ciência e Tecnologia de Rondônia, Colorado do Oeste, Brazil, (4)CBMEG - UNICAMP, Campinas, Brazil

Uncoupling Proteins (UCPs) are inner mitochondrial membrane proteins that uncouple substrate oxidation in mitochondria from ATP synthesis by the respiratory chain. These proteins are crucial for protection against reactive oxygen species (ROS) and maintenance of mitochondrial homeostasis. Studies using single T-DNA insertion mutants for the three *UCP* genes present in *Arabidopsis thaliana* (*AtUCP1-3*) revealed a reduced fertility phenotype (i.e. short silique length and seed yield) in both *atucp1* and *atucp2*. Reciprocal crosses between the investigated mutants and Col-0 uncovered defects in male (*atucp1* and *atucp2*) and female (*atucp1*) gametophytes. This was further substantiated by the down-regulation in flowers from *atucp1-2* at different developmental stages of transcription factors implicated in female (*AtHEC1-3*) and male (*AtDYT* and *AtAMS*) gametophyte development, respectively. Also in agreement and confirming the spatial association of *AtUCP1-2* expression with reproductive organ/tissues, transgenic tobacco flowers harboring *AtUCP1* or *AtUCP2* promoter:GUS fusions displayed reporter activity mainly in the stigma and anthers. Interestingly, the mutant flowers exhibited high levels of ROS compared to Col-0, suggesting that the reduced fertility phenotype might be a result of unbalanced ROS accumulation in reproductive organs. In addition, a significant delay in flowering time and a reduction in seed and biomass production was observed in an *atucp1/atucp2* double-mutant. Taken together, these data give support for an important role of UCPs in flower metabolism and overall plant fertility.

PE0578: Brassicas, *Arabidopsis*, and related

Genome-Wide Association Studies Provide Insights into Rapeseed Tolerance to Flooding during Germination

Jijun Li, Huazhong Agricultural University, Wuhan, China

Rapeseed is an important oil crop in the world and flooding severely affects the seed emergence and germination after sowing. In order to explore the genetic and molecular mechanisms of rapeseed tolerance to flooding during germination, 9 phenotypic traits and membership function evaluate waterlogging (MFVW) were measured for 505 inbred lines of rapeseed at the germination stage. Genome-wide association studies (GWAS) were performed using a mixed linear model to dissect the genetic basis of rapeseed tolerance to flooding during seed germination. We identified 78 SNP loci significantly associated with flooding tolerance ($-\log(P) > 7.0$) in rapeseed. Five of the 78 SNP loci were identified under MFVW and more than one other trait. We identified a number of tolerant and sensitive germplasms based on MFVW. Transcriptome analysis suggested flooding stress changed cell wall organization and oxygen-containing compound response in both R and S genotypes. Genes found to be specific to the R genotype related to polysaccharide metabolic process and organonitrogen compound. Forty genes were identified in both the SNP-trait association and transcriptome sequencing analyses, including oxidases, membrane-associated protein and plant transcription factors.

PO0579: Brassicas, Arabidopsis, and related

Characterisation of DNA Methylation Status in *Brassica napus* in Response to *Leptosphaeria maculans*

Soodeh Tirnaz, Clementine Merce, Anita Severn-Ellis, Philipp E. Bayer, David Edwards and Jacqueline Batley, University of Western Australia, Perth, WA, Australia

Canola (*Brassica napus*), the most important species of the Brassicaceae family, is grown worldwide for its seeds which are crushed for the edible oil. Canola production is threatened by diseases such as blackleg, caused by the fungal pathogen *Leptosphaeria maculans*, which can cause up to 100% yield loss. The resistance response highly depends on the expression regulation of defence genes. Epigenetic marks, such as DNA methylation, are used by plants to regulate gene expression. Pathogen-induced DNA methylation modification is an important regulatory mechanism in a plant's defence system. In this work, the DNA methylation pattern of defence genes was investigated within resistant and susceptible cultivars of *B. napus* in response to *L. maculans*. Leaves from resistant and susceptible cultivars were harvested 9 days after inoculation and their methylation pattern was studied using a SeqCap Epi probe set. Comparative analysis was conducted between the control and infected samples, between the cultivars and between leaf samples. The results revealed the majority of differentially methylated genes between the resistant and susceptible cultivars were involved in Biological Process, Cellular Component and Molecular Function; among them more than 200 were involved in defence mechanisms. The comparative analysis among leaves show the majority of defence genes are hypo- or hyper-methylated in the first and second true leaves, but not in the third true leaf. It is expected that the outcome will assist in understanding the role of DNA methylation in plant defence against disease and ultimately help Brassica breeders to better improve resistance against blackleg disease.

Key words: Brassicaceae, *Brassica napus*, Blackleg, DNA methylation, Resistance gene

PE0580: Brassicas, Arabidopsis, and related

Eight High-Quality Genomes Reveal Pan-Genome Architecture and Ecotype Differentiation of *Brassica napus*

Jiaming Song, Zhiling Guan, Jianlin Hu, Chaocheng Guo, Zhiqun Yang, Shuo Wang, Dongxu Liu, Bo Wang, Shaoping Lu, Run Zhou, Wenzhao Xie, Yuanfang Cheng, Yuting Zhang, Kede Liu, Qingyong Yang, Ling-Ling Chen and Liang Guo, Huazhong Agricultural University, Wuhan, China

Rapeseed (*B. napus*) is the second most important oilseed crop in the world, but the genetic diversity underlying its massive phenotypic variations remains largely unexplored. Here, we report the sequencing, *de novo* assembly and annotation of eight *B. napus* accessions. Using pan-genome comparative analysis, millions of small variations and 77.2–149.6 Mb presence/absence variations (PAVs) were identified, and >9.4% of the genes contained large-effect

mutations or structural variations. PAV-based genome-wide association study (PAV-GWAS) directly identified causal structural variations for silique length, seed weight and flowering time in a nested association mapping population with ZS11 (reference line) as the donor, which were not detected by single-nucleotide polymorphisms based GWAS (SNP-GWAS), demonstrating that PAV-GWAS was complementary to SNP-GWAS in identifying associations to traits. Further analysis showed that PAVs in three *FLOWERING LOCUS C* genes were closely related to flowering time and ecotype differentiation. This study provides resources to support a better understanding of the genome architecture and acceleration of the genetic improvement of *B. napus*.

PO0581: Brassicas, Arabidopsis, and related

Dissection of the Genetic Control of Plant Architecture by Associative Transcriptomics in *Brassica napus*

Meili Xie¹, Chaobo Tong², Yuanyuan Zhang³ and Shengyi Liu², (1)Oil Crops Research Institute of Chinese Academy of Agricultural Sciences, Wuhan, China, (2)Oil Crops Research Institute of Chinese Academy of Agricultural Sciences, Wuhan, China, (3)Oil Crops Research Institute, Chinese Academy of Agricultural Sciences, Wuhan, China

Plant architecture refers to the morphological characteristics and spatial arrangement of the plant crops and has shown significant impacts on crop adaptation and yield potential. Although few genes related to plant architecture have been identified, little is known about the genetic control of the important traits in *Brassica napus*. Associative Transcriptomics is an advanced method to identify molecular markers associated with trait variation across a diversity panel, exploiting both gene sequence variation and gene expression variation. Here a genetic diversity panel of more than 300 accessions are chose to dissect the genetic control of plant architecture. Using associative transcriptomics, we identify molecular markers for a suite of important traits. Furthermore, based on conserved synteny with *Arabidopsis thaliana* we are able to provide a biological context to the candidate genes detected and verify the function of these genes in *Arabidopsis*. This study not only furthers our understanding of the gene network for plant architecture but also provides a set of markers for the breeding programmes.

PE0582: Brassicas, Arabidopsis, and related

Combining GWAS and High-Throughput Phenotypic Platform Dissect Genetic Architecture of Salt Tolerance in *Brassica napus*

Guofang Zhang, Huazhong Agricultural University, Wuhan, China

Adverse environmental factors such as drought, low temperature, heat, salinity and so on inhibit crop growth and development. Although researchers have made a lot of achievement of understanding plant salinity stress tolerance, its regulatory pathways remain to be explored. As the soil salinization is becoming more and more severe, it is pivotal to excellent salinization tolerant cultivars for stable crop production. *Brassica napus* is vital oil crop in the world, but it is facing severe soil salinization. Therefore, it is important for us to understand salinity adaptability and explore additional genetic resources for the salinity tolerance improvement of *Brassica napus*. We investigated the genetic variation of 505 *Brassica napus* accessions under low and high salt concentration stress conditions by genome-wide association studies (GWAS). We took full use of the high throughput phenotyping platform to phenotype the natural variation of 505 *Brassica napus* accessions under two different salt stress conditions . We have identified dozens of important visible light traits. Several salt-tolerant and -sensitive accessions have been identified, which can be used as breeding materials. GWAS was performed to detect the genetic loci for salt relevant traits. So far, we have identified several loci related to salt stress and candidate genes were predicted for functional studies. The significant salt tolerance-related loci and candidate genes detected in this study can help the marker-assisted selection of improved *Brassica napus* to cope with soil salinization stress.

PO0583: Brassicas, Arabidopsis, and related

Genome-Wide Association Study Identifies Loci Associated with Winter Hardiness in Canola (*Brassica napus* L.)

Wun S. Chao¹, James V. Anderson¹, David P. Horvath¹ and Michael J. Stamm², (1)USDA-ARS, Fargo, ND, (2)Kansas State University, Manhattan, KS

Canola is an important oilseed crop contributing to the global demand for oil production. Winter canola generally produces greater yields than spring canola. However, yearly winter canola acreage is limited due to its inability to withstand the harsh winter conditions that occur in many northern regions of the U.S.A. To identify loci associated for freezing tolerance in canola, we conducted a genome-wide association study (GWAS) using a genotyped diversity panel containing 399 accessions consisting primarily of winter canola. Genotyping-by-sequencing (GBS) of this canola diversity panel population identified 251,575 single-nucleotide polymorphisms (SNPs) with an average distance of 3377.3 bases; more than three quarters of the SNPs are 1877 bases apart or closer. Cluster analysis confirmed a high degree of diversity within the population of accessions, with minimal sub-clustering of genotypes. One-month-old plants from this population were acclimated (8 weeks at 5°C) and then subjected to a freezing treatment (-15°C for 4 hours) before they were phenotyped for freezing survival and chlorophyll fluorescence (Fv/Fo) in a greenhouse. There was reasonable correlation observed between visual damage and chlorophyll fluorescence ratings among the top associated loci. The resulting numerical values for phenotypes were used for association analyses with the identified SNPs. Thirty-two significant loci were identified for the phenotypes scored, with several showing significance for multiple phenotypes. Thirteen candidate genes were identified as previously associated with freezing tolerance and photosynthesis. Several candidate genes underlying identified loci were also differentially regulated by cold.

PE0584: Brassicas, Arabidopsis, and related

Identification of New QTLs for Resistance to *Plasmodiophora brassicae* in *Brassica napus* Using Genome Wide Association Mapping

Abdulsalam Dakouri¹, Mebarek Lamara², Masud Karim¹, Bruce G. Gossen¹, Gary Peng¹, Sue Boyetchko¹ and Fengqun Yu¹, (1)Agriculture and Agri-Food Canada, Saskatoon, SK, Canada, (2)Université du Québec en Abitibi-Témiscamingue, Rouyn-Noranda, QC, Canada

Clubroot of canola (*Brassica napus*), caused by the obligate pathogen *Plasmodiophora brassicae* Woronin, is a major disease worldwide. Genetic resistance remains the best strategy to manage this disease. The objective of the study was to identify and map new sources of resistance to clubroot in *B. napus* using genome-wide association mapping. The reaction of a collection of 177 accessions to four highly virulent pathotypes of *P. brassicae* was assessed. These pathotypes were selected because they were most recently identified and showed different virulence patterns on the Canadian clubroot differential (CCD) lines. The collection was then genotyped using genotyping by sequencing (GBS) method. Multi-locus mixed linear model (MLMM) was used to perform the association analysis. The majority of accessions were highly susceptible (70–100 DSI), while few individual accessions showed strong resistance (0–20 DSI) to 5X (2 accessions), 2B (7 accessions), 3A (8 accessions) and 3D (15 accessions). In total, 301,753 SNPs were mapped to 19 chromosomes. Population structure analysis indicated that the 177 accessions belong to two major populations. SNPs were associated with resistance to each pathotype using MLMM. In total, 23 significant SNP loci were identified, with 14 SNPs mapped to the A-genome and 9 to the C-genome. The SNPs were associated with resistance to pathotypes 5X (4 SNPs), 2B (9), 3A (5) and 3D (5). A blast search of 2 Mb upstream and downstream identified 61 disease resistance genes, of which 24 belonged to TIR-NBS-LRR proteins and 20 belonged to CC-NBS-LRR proteins. The distance between a SNP locus and the nearest resistance genes ranged from 0.11–1.66 Mb. This indicated that NBS-LRR gene family might have an important role in clubroot resistance in *B. napus*. The resistant *B. napus* lines and the SNP markers identified in this study can be used for breeding for resistance to clubroot and contribute to understanding the genetic mechanism of resistance to clubroot.

P00585: Brassicas, Arabidopsis, and related

Improved *Brassica oleracea* JZS Reference Genome Uncovers That Diversification Among Subspecies Was Driven By LTR-RT

Xu Cai¹, **Jian Wu¹**, Jianli Liang¹, Limei Yang¹, Xiaowu Wang¹ and , (1)Institute of Vegetables and Flowers, Chinese Academy of Agricultural Sciences, Beijing, China

Brassica oleracea is an important vegetable crop that has provided ancestor genomes of the two most important *Brassica* oil crops, *Brassica napus* and *Brassica carinata*. The current *B. oleracea* reference genome (JZS, also

named 02-12) displays problems of large mis-assemblies, low sequence continuity, and low assembly integrity, thus limiting genomic analysis. We developed a reliable misjoins correction pipeline and report an updated assembly of the *B. oleracea* reference genome (JZS_v2) derived using single-molecule sequencing and chromosome conformation capture technologies. We assembled an additional 83.16 Mb of genomic sequences, and the updated genome features a contig N50 size of 2.37 Mb, representing an ~84-fold improvement. We detected a new round of long terminal repeat retrotransposon (LTR-RT) burst in the new assembly. Comparative analysis with the reported genome sequences of two other subspecies of *B. oleracea* (TO1000 and HDEM) identified extensive gene order and gene structural variation, and approximately half of the presence/absence variation (PAV) genes were due to gene fractionation after sub-speciation. We further found that different amplification patterns of *athila*, *tat*, and *Del* Gypsy subgroups were a factor in the diversity of *B. oleracea* subspecies, and that LTR-RTs could directly influence genes with specific functions to drive diversification among subspecies. We provide a high-quality *B. oleracea* reference genome, and we found extensive gene structural variation and LTR-RT-driven diversification among subspecies.

PE0586: Brassicas, Arabidopsis, and related

Complete Chloroplast Genome Sequence and Variation Analysis of *Brassica oleracea* L.

Yuanyuan Xu, Institute of Vegetable Crops, Jiangsu Academy of Agricultural Sciences, Nanjing, China

Abstract:

Cabbage (*Brassica oleracea* L. var *capitata*) is one of the most important vegetable crops of the Cruciferae family, which is widely distributed throughout China. However, studies on the chloroplast genome of this important agricultural species are conspicuous by their absence. In this study, the complete chloroplast (cp) genomes of *Brassica oleracea* L. including JF and JF-CMS were sequenced, analyzed, and compared to four other species of the Cruciferous family. Whole sequence alignment was also carried out followed by phylogenetic analysis. The size of the JF chloroplast genome was found to be 153,363 bp including one large single copy (LSC) region of 83,136 bp and one small single copy (SSC) region of 17,833 bp, separated by two inverted repeat (IR) regions of 26,197 bp each. The GC content of the whole genome was 36.36%, while those of the LSC, SSC and IR were 29.10%, 34.15%, and 42.34%, respectively. A total of 134 genes were identified, of which 87 were protein-coding genes, 39 were transfer RNAs, and eight were ribosomal RNAs. In the repeat structure analysis, a total of 49 long repeats and 271 simple sequence repeats (SSRs) were discovered in the JF chloroplast genome. The existence of a large amount of SSRs in the genome carries the potential for future research on population genetics. Whole genome comparison results revealed that LSC and IR regions were more divergent than SSC regions and that a higher divergence was observed in the noncoding than in the coding regions. The phylogenetic relationship of 38 chloroplast genomes (represented by 40 sequences) revealed that JF and JF-CMS were not part of a single clade. This study uncovered the unique characteristics of the cabbage chloroplast genome, which will have potential applications in species identification and agricultural research, and also inform research on breeding cabbage for genetic improvement.

PO0587: Brassicas, Arabidopsis, and related

Genome Assembly of Five Different *Brassica oleracea* Morphotypes

Chengcheng Cai, Johan Bucher, Fernanda A. de F. Guedes, Richard Finkers and Guusje Bonnema, Wageningen University & Research, Plant Breeding, Wageningen, Netherlands

Brassica oleracea (2n=18) is an economically important crop which is renowned for its outstanding morphological diversity, including broccoli, cauliflower, collard green, heading cabbage, kale, kohlrabi, ornamental, brussels sprouts, tronchuda, etc. The publicly available reference genomes are not sufficient to represent genome characteristics for all *B. oleracea* morphotypes due to the existence of structural variations. Here we report *de novo* high-quality, chromosome-scale genome assemblies of five representative *B. oleracea* morphotypes (broccoli, cauliflower, kohlrabi, kale and white cabbage) using a combination of long-read sequencing, short-read sequencing and Bionano optical mapping technologies. All of the five assemblies show high contig N50 of >6Mb, as well as high sequence accuracy of 99.99% after polishing. Finished Bionano hybrid scaffolding for three morphotypes shows that ~98% of contig sequences can be anchored into ~40 scaffolds with scaffold N50 of >30Mb. These three assemblies contain scaffolds which represent entire chromosome arms. Our study provides five high quality

reference genomes for the Brassica community, representing valuable resources for structural variation studies within species as well as pangenome construction of *B. oleracea*.

PE0588: Brassicas, Arabidopsis, and related

Developing Genomic Resource in Watercress (*Nasturtium officinale*, R.)-Exploring G X E Interaction for Plant Morphology and Anti-Cancer Properties

Nikol Voutsina¹, Yufei Qian², Lauren E. Hibbert¹, Annabelle Damerum³, Baohua Li², Daniel Kliebenstien², Graham Packham¹, Mark J. Coldwell¹, Mark A. Chapman¹, Tingting Zhu⁴, Ming-Cheng Luo³ and Gail Taylor³, (1)University of Southampton, Southampton, United Kingdom, (2)University of California, Davis, Davis, CA, (3)University of California, Davis, CA, (4)Department of Plant Sciences, University of California, Davis, Davis, CA

Watercress (*Nasturtium officinale*, R.) is a highly nutritious leafy green vegetable that belongs to the mustard family and is ideal for indoor vertical agricultural systems, since naturally, it grows in an aquatic or semi-aquatic habitat. The peppery taste of watercress is the indicator of mustard oils which help reduce inflammation and chronic damage in cells and has been shown to have a role in cancer prevention *in vitro*.

Watercress has a rich health-related phytonutrient profile including secondary plant metabolites antioxidants and glucosinolates, especially gluconasturtiin and phenethyl isothiocyanate. We describe the development of the first 259 F₂ progeny mapping population, the use of reduced representation Genotyping-by-sequencing (GBS) for marker discovery and the construction of the first genetic linkage map with a 1,733 cM length and a mean inter-marker distance of 5.8 cM. We have identified seventeen novel quantitative trait loci (QTL) for plant morphology and nutritional composition.

We are exploring the effects of G x E interactions on phenotypic traits for a set of 'extreme' F₃ lines, including growth, morphology, antioxidant capacity, glucosinolates and sugar content, produced in contrasting field locations. After identifying and grouping the "extreme" lines these were bulked and transcriptome analysis performed to identify candidate genes for further analysis. The long-term aim is to use genomics-assisted breeding to develop more nutritious and better anti-cancer watercress with high resource-use efficiency.

PO0589: Brassicas, Arabidopsis, and related

Cytological and Molecular Studies on a Male Sterile Variant Y003-21-8 and Its Fertile Revertant in Yellow Mustard (*Sinapis alba*)

Fangqin Zeng and Bifang Cheng, Agriculture and Agri-Food Canada, Saskatoon, SK, Canada

A male sterile (ms) variant Y003-21-8 was discovered in yellow mustard (*Sinapis alba*). It was crossed as female with 12 different inbred lines with the aim to identify its restorer and maintainer lines. All the tested lines functioned as a restorer for Y003-21-8. Therefore, Y003-21-8 was maintained by cloning in the greenhouse. Interestingly, some of the clones produced branches that were male fertile (mf) revertant. Scanning electron microscopy analysis indicated that both the sterile and fertile pollen grains had a reticulate exine layer. However, transmission electron microscopy studies revealed that the nexine and intine layers were present in the microspore of the mf revertant, but absent in that of the ms variant. This suggested that the genes controlling the nexine and intine formation might not function normally and thus result in the microspore abortion in the ms variant. RNA-Seq and differential gene expression analysis revealed that the two genes, *At4g14080* being involved in pollen wall development and *TRANSPOSABLE ELEMENT SILENCING VIA AT-HOOK (TEK)* controlling nexine formation, were down-regulated in the ms variant, but up-regulated in the mf revertant. Additionally, the *putative caffeoyl-CoA O-methyltransferase At1g67980* was up-regulated in the ms variant, but down-regulated in the mf revertant. The cloned cDNAs of *At4g14080* and *TEK* from the ms variant were identical in size and nucleotide sequence with those of the mf revertant, respectively. Investigation on the molecular mechanism underpinning the sterility of the ms variant and the fertility of the mf revertant is in progress.

PE0590: Brassicas, Arabidopsis, and related

Transcriptome Analysis Before and After Salicylic Acid Treatments in *Brassica rapa* L.

Naomi Miyaji, Takeshi Takasaki-Yasuda and Ryo Fujimoto, Graduate School of Agricultural Science, Kobe University, Kobe, Japan

Salicylic acid (SA) is a plant hormone, which plays an important role in regulation of the defense signaling network. In *Brassica rapa* L., the molecular mechanism of SA response is unclear, and SA responsive genes have not been identified at the whole genome level. In this study, transcriptome analysis before and after SA treatments in two cultivars of *B. rapa* was performed. Seven day old seedlings grown on MS medium without SA were transplanted to MS medium with SA, and cotyledons were harvested for RNA extraction before (0 h) and at 72 h after SA treatment. Sequenced reads were mapped to the *B. rapa* reference genome version 1.5 by hisat2, and differentially expressed genes (DEGs) between samples with and without SA treatments were identified by cuffdiff. 550 and 597 genes were up-regulated after SA treatment in each cultivar, and 265 genes overlapped. 389 and 655 genes were down-regulated after SA treatment in each cultivar, and 225 genes overlapped. In *Arabidopsis thaliana*, 217 genes have been identified as SA-induced genes (SAIGs) (Blanco et al. 2009). In total, 422 genes in *B. rapa* are orthologs of SAIGs in *A. thaliana*. 42 and 44 of 422 genes were up-regulated by SA treatment in each cultivar of *B. rapa*, and 24 genes overlapped. Of the 24 DEGs in *B. rapa*, some *WRKY* and *glutathione S-transferase (GST)* genes are included.

PE0592: Brassicas, Arabidopsis, and related

A Comprehensive Genetic Linkage Map Identified QTL Associated with Flowering Behaviour in Winter *Camelina sativa*

Raju Chaudhary, University of Saskatchewan, Saskatoon, SK, Canada and Isobel Parkin, Agriculture and Agri-Food Canada, Saskatoon, SK, Canada

Camelina sativa is an oilseed crop that is gaining attention due to its agronomic and seed quality traits, with a balanced omega fatty acid profile, optimal for human and animal consumption. The rarer winter form of *C. sativa* has production and environmental advantages. Interspecific hybridization between related *Camelina* species was utilized to identify quantitative trait loci (QTL) associated with flowering behaviour in winter *C. sativa*. Spring *C. sativa* was crossed with two different wild relatives (*C. alyssum* and *C. sativa ssp. pilosa*) as a source of winter habit to generate genetic linkage maps. Two F₂ population, one from *C. sativa* X *C. alyssum* (Csa: 169 lines) and one from *C. sativa* X *C. sativa ssp. pilosa* (Csp: 118 lines) were studied. In the Csa population, 3634 genotyping-by-sequencing (GBS) based markers were mapped to 20 chromosomes of the reference *C. sativa* genome which accounted for 3052.33 cM of mapping distance; likewise, from the Csp population, 1734 GBS markers were mapped, corresponding to a map distance of 2646.95 cM. This study identified two major and two minor QTLs associated with flowering behaviour in *C. sativa*. Both major QTLs were found in proximity to FLOWERING LOCUS C (FLC); however, they were mapped to different homologous subgenomes in the two populations. In addition, genetic linkage maps suggested minor improvements to the genome assembly on Chr16. Further work is underway to characterize the region underlying the QTLs, which will help to dissect the flowering mechanism in *Camelina*, and generate tools to facilitate the winter *Camelina* breeding program.

PO0593: Brassicas, Arabidopsis, and related

Nanopore Assembly of Two *Brassica nigra* Genomes Provides Basis for Genome Completeness and Comprehensive Transposon Dynamics

Sampath Perumal¹, Chu Shin Koh², Lingling Jin³, Steve Robinson⁴, Erin Higgins⁵, Zahra-Katy Navabi⁶, Miles Buchwaldt⁵, Andrew G. Sharpe² and Isobel Parkin⁵, (1)Agriculture and Agri-Food Canada, SASKATOON, SK, Canada, (2)Global Institute for Food Security (U of S), Saskatoon, SK, Canada, (3)Thompson Rivers University, Kamloops, BC, Canada, (4)Agriculture and Agri-Food Canada, Saskatoon, SK, Canada, (5)Agriculture and Agri-Food Canada, Saskatoon, SK, Canada, (6)Agriculture and Agri-Food Canada, Saskatoon, SK, Canada

Continuing improvements in long-read sequencing technology have provided a wealth of opportunity for genome assembly and analyses, providing greater genome coverage and a more complete picture of complex sequence architecture that proved inaccessible with short read data. Taking advantage of these developments, high-quality

reference assemblies were generated for two *Brassica nigra* genomes (Ni100 and CN115125) which help to understand genome structure and evolution in the relatively neglected Brassica B genome. The N50 values of the contigs for the two assemblies ranged from 0.288 Mb (963 contigs) to 17.1 Mb (58 contigs) reflecting rapid improvements in the technology over the course of the project. A de novo short read assembly for one of the two lines corroborated genome integrity and allowed a clear determination of sequence related error rates (0.002%). The contiguity and coverage allowed unprecedented access to low complexity regions of the genome, including pericentromeric regions, and coincidence of hypo-methylation, identified both directly from long read data and through traditional bisulfite sequencing, suggested at least partial capture of the active centromeres for each chromosome. A comprehensive study of transposon dynamics was possible, identifying a novel centromeric associated Ale class I element which appears to have proliferated through relatively recent nested transposition events (<1 million years ago). The assembly of a mesopolyploid with multiple levels of genome duplication using cost-effective ONT long-read sequencing revolutionizes genome sequencing in polyploid genomes, and will allow the development of important resources for the functional and comparative analyses of Brassica crops and their related species.

PE0594: Brassicas, Arabidopsis, and related

The Radish Genome Database (RadishGD): An Integrated Information Resource for Radish Genomics

Hee-Ju Yu¹, Seunghoon Baek², Ara Cho² and **Jeong-Hwan Mun²**, (1)The Catholic University of Korea, Bucheon, South Korea, (2)Myongji University, Yongin, South Korea

Radish (*Raphanus sativus* L.) is an important root vegetable crop in the family Brassicaceae, which provides diverse nutrients for human health and is closely related to the *Brassica* crop species. Recently, we sequenced and assembled the radish genome into nine chromosome pseudomolecules. In addition, we developed diverse genomic resources, including genetic maps, molecular markers, transcriptome, genome-wide methylation, and variome data. In this study, we describe the radish genome database (RadishGD), including details of datasets that we generated and the web interface that allows access to these data. RadishGD comprises six major units that enable researchers and general users to search, browse, and analyze the radish genomic data in an integrated manner. The Search unit provides gene structures and sequences for gene models through keyword or BLAST searches. The Genome browser displays graphic representations of gene models, mRNAs, repetitive sequences, genome-wide methylation, and variomes among various genotypes. The Functional annotation unit offers gene ontology, plant ontology, pathway, and gene family information for gene models. The Genetic map unit provides information about markers and their genetic locations using two types of genetic maps. The Expression unit presents transcriptional characteristics and methylation levels for each gene in 18 tissues. All sequence data incorporated into RadishGD can be downloaded from the Data resources unit. Efforts to improve RadishGD by adding additional data and analyses are ongoing. We are currently studying radish genome sequencing and assembly based on long-read sequencing technology to improve the genome assembly. In parallel, we are also developing phenotype data and whole-genome resequencing data for the radish core collection consisting of 125 accessions. RadishGD will be continually updated to serve as a community resource for radish genomics and breeding research.

PO0595: Brassicas, Arabidopsis, and related

Genome Sequencing of the Ethiopian Mustard (*Brassica carinata*): Investigation of the Genetic Evolution and Diversity for Agronomic Traits

Won Yim, Mia L. Swain, David Curdie, John C. Cushman and Jeff F. Harper, University of Nevada - Reno, Reno, NV

The Ethiopian mustard (*Brassica carinata*) is an important oil crop in Africa which include BBCC subgenome among the three allopolyploid genome. To improve our understanding of molecular genetic basis of the evolutionary relationships among these Brassica species, we assembled an allopolyploid *Brassica carinata* genome by single-molecule reads and chromatin conformation map integrated to genomic and genetic map. We selected an advanced generation inbred line of *B. carinata* var. Gomenzer for whole-genome sequencing. We estimated the size of the Gomenzer genome at 1.2 Gb by flow cytometry. The genome assembly was generated for *B. carinata* using two different genome sequencing platform, PacBio and HiC and obtained in a depth of coverage of 124X and 30X,

respectively. The genome was assembled by trio-binning method and we discovered that the subgenome B and C had an independent evolutionary history. Also, the genome was indicated thousands of fractionation *via* exotic introgression and induced novel variation. Identification of allopolyploid genome assembly creates a rich genetics ‘parts list’ resource for ongoing brassica crop improvement and provide more information and insights to understand the genomic variation.

PE0596: Cotton

Analysis of Categorical Traits in the National Germplasm Collections. Case Study: Pima and Upland Cotton

Daniel Restrepo-Montoya¹, Amanda M. Hulse-Kemp², Lori L. Hinze³, Jodi A. Scheffler⁴, Janna Love⁵, Richard G. Percy³, Don C. Jones⁶, Candace H Haigler¹ and James Frelichowski⁵, (1)North Carolina State University, Raleigh, NC, (2)USDA-ARS-GBRU, Raleigh, NC, (3)USDA-ARS, Southern Plains Agricultural Research Center, College Station, TX, (4)USDA-ARS, Stoneville, MS, (5)USDA-ARS, College Station, TX, (6)Cotton Incorporated, Cary, NC

Qualitative morphological traits are relatively stable across environments and are commonly used to evaluate genetic diversity. In the past two decades, molecular markers have largely superseded phenotypic traits in diversity surveys. However, qualitatively inherited traits continue to be used as sources for: a) descriptors in cataloging accessions of germplasm collections; and b) germplasm registration (Plant Variety Protection, patent, and/or publication). Recently, the USDA-ARS Crop Germplasm Research Unit in College Station, Texas updated previous disparate descriptor schemes with a standardized rating scale that encompasses the diversity observed across *Gossypium* species. A large portion of the U.S. National Cotton Germplasm Collection has been evaluated under this standardized scheme for the last nine years. The current research focuses on comparative analysis of morphological traits within three major groups of *Gossypium* accessions, including Upland cotton (*Gossypium hirsutum* cultivars and landraces) and Pima cotton (*Gossypium barbadense*). Using the standardized scores of 36 traits/descriptors (e.g. leaf hairs, boll nectaries, and seed type), several analyses were performed including categorical data distributions, bivariate associations, and clustering across accessions. The categorical distributions illustrate both intra- and interspecies phenotypic diversity. The association test suggests a significant bivariate association among some of the 36 categorical descriptors evaluated, and the encoded transformation of the categorical variables allowed evaluation of the phenotypic descriptors by clustering analysis. The results provide novel insights for researchers interested in using the germplasm collection and for curators maintaining the collection. Broadly, the study exemplifies the use of computational strategies to extract additional information from databases of germplasm collections.

PO0597: Cotton

Genetic Dissection of an Allotetraploid Interspecific CSSLs Guides Interspecific Genetics and Breeding in Cotton

De Zhu, National Key Laboratory of Crop Genetic Improvement, College of Plant Sciences & Technology, Huazhong Agricultural University, Wuhan 430070, Hubei, China, wuhan, China

The low genetic diversity of Upland cotton limits the potential for genetic improvement. Making full use of the genetic resources of Sea-island cotton will facilitate genetic improvement of widely cultivated Upland cotton varieties. Here, we revealed the genetic effects in an interspecific species using chromosome segments substitution lines (CSSLs) between *Gossypium hirsutum* (Gh) × *G. barbadense* (Gb). By whole genome re-sequencing, 11,653,661 high-quality single nucleotide polymorphisms (SNPs) were identified which ultimately constructed 1,211 recombination chromosome introgression segments from Gb. The sequencing-based physical map provided more accurate introgressions than the PCR-based markers. By exploiting CSSLs with mutant morphological traits, the genes responding for leaf shape, leaf size and fuzz-less mutation in the Gb were identified. Based on a high-resolution recombination bin map to uncover genetic loci determining the phenotypic variance between Gh and Gb, 64 quantitative trait loci (QTL) were identified for 14 agronomic traits. Outbreeding depression in the CSSLs revealed that the minor effect of the Gb alleles was diluted in the Gh tank background. We proposed that different Gb alleles should be combined with the beneficial Gh alleles to improve the fiber quality of Gh cultivars. Surprisingly, multiple alleles of Gb showed extremely high value in enhancing cottonseed oil content (SOC), which is seldom studied and discovered before; and one candidate gene, *Gbar_A01G002860.1*, encoding a pyruvate

dehydrogenase kinase, was identified for regulating the metabolism of tricarboxylic acid in the mitochondria, which plays a pivotal role in the fatty acid biosynthesis. Collectively, our results provide insights into the genetic effects of Gb alleles on the Gh, and provide guidance for the utilization of Gb alleles in interspecific breeding.

PE0598: Cotton

The Complexity of Pentatricopeptide Repeat Protein in a Rf-Gene Region Uncovered a Special Evolution Model in Cotton

Gao Bin, National Key Laboratory of Crop Genetic Improvement, College of Plant Sciences & Technology, Huazhong Agricultural University, Wuhan, Wuhan, China

Hybrid cotton production is a common way to utilize heterosis, and ‘Three Line’ system plays a crucial part in cotton breeding work. In the past decades, several sets of ‘Three Line’ have been developed and some with a strong restore line like *G. harknessii*-type CMS have been used in breeding, but the molecular mechanism of cotton ‘Three Line’ is limited to the mapping of restore gene (*Rf*) delimiting in a PPR-cluster region. In our study, the CMS line with *G. hirsutum* cytoplasm named as 6001A and the restore line named as 7R13 was used to construct F₂ mapping population, and we mapped the *Rf* gene on D05 chromosome according to BSA-seq, followed by fine mapping the region to 1.9 Mb. To check the variation of the 17 PPR like sequences predicted in the region, nested PCR was used to amplify all the sequences, and four variant PPRs were discovered with SNP, Indel or duplicated PPR motifs. Based on the variant PPR sequence, we developed a CAPS marker which can be used to distinguish 7R13 from other materials easily.

Furthermore, we developed a set of primers with special adapters covering conserved regions of PPR sequences, and PCR products amplified from cDNA of different development periods were submitted to PCR-free library construction followed by high-throughput sequencing; as a result, we detected many diverse PPR fragments and relatively different read-counts PPR fragments between 6001A and 7R13, indicating different expression level of special PPR genes. Then, we re-sequenced 7R13, interestingly, we identified a cluster of PPR genes, with more than three PPR genes arrayed, and more than 4 new PPR genes different from previously discovered were detected. The results revealed that the evolution of restore gene has experienced a rapid expansion of PPR genes in the D05 region.

PO0599: Cotton

CottonGen, a Central Platform to Facilitate Knowledge Discovery in Cotton Research

Jing Yu¹, Sook Jung¹, Chun-Huai Cheng¹, Taein Lee¹, Ping Zheng¹, Katheryn Buble², Jodi L. Humann¹, James Crabb³, Heidi Hough¹, Don C. Jones⁴, B. Todd Campbell⁵, Josh Udall⁶ and Dorrie Main³, (1)Washington State University, Pullman, WA, (2)Washington State University, Pullman, WA, (3)Department of Horticulture, Washington State University, Pullman, WA, (4)Cotton Incorporated, Cary, NC, (5)USDA-ARS, Florence, SC, (6)USDA-ARS, College Station, TX

CottonGen (<https://www.cottongen.org>) integrated significantly more data in 2019: New whole genome assembly data include those from 5 diploids (*G. arboreum*, *G. raimondii*, *G. turneri*, *G. austral*, *Gossypioideis kirkii*) and 5 tetraploids (*G. hirsutum*, *G. barbadense*, *G. mustelinum*). Cotton sequences, curated from NCBI, has also been updated and now contains over 53K new genes and over 205K new proteins. Seven new (*G. arboreum*, *G. barbadense*, *G. turneri*, *G. kirkii*, *G. austral*, *G. mustelinum*) or updated (*G. hirsutum*) pathways are available from CottonCyc database. 282 cotton trait descriptors, developed from QTL publications and evaluation data, have been submitted to Crop Ontology (<http://www.cropontology.org>). Other new data added in 2019 include 8000 non-fiber traits and 5,000 new fiber trait data from RBTN, 132 QTLs, 280 QTL trait data, 2,230 SSR markers collected from peer-viewed publications. New and updated tools include the interactive and customizable search tool, MegaSearch and new functionalities including SNP genotype data management in BIMS.

PE0600: Cotton

Using Exotic *Gossypium hirsutum* to Increase Genetic Variation in Elite *G. hirsutum* Varieties, DES56 and Acala Maxxa

Ellen Regina E Skelton¹, Tariq Shehzad¹, Susan A. Auckland¹, Lisa K. Rainville¹, Jon S. Robertson², Don C. Jones³, Changying Li⁴ and Andrew H. Paterson¹, (1)Plant Genome Mapping Laboratory, University of Georgia, Athens, GA, (2)University of Georgia, Athens, GA, (3)Cotton Incorporated, Cary, NC, (4)College of Engineering, University of Georgia, Athens, GA

The drastically reduced genetic variation in crops due to repeated genetic bottlenecks including domestication, human dispersal, selective breeding, and transformability may be combated by introducing variation from wild relatives. To introduce more variation into the *Gossypium hirsutum* genome, we hybridized wild *G. hirsutum* lines with two elite lines, DES56 and Acala Maxxa. These hybrids were selfed to the F6 generation, which resulted in 24 QTL mapping populations. Ten lines of each population were randomly chosen, grown, and phenotyped in 2015 and 2016, which represent the F5 and F6 generations. 2016 had two replicates of each line, laid out in a randomized complete block design. DNA was extracted from each line for genetic mapping using PCR and gel electrophoresis with acrylamide gels. Data were collected on fiber yield and quality to compare if the populations differed from each other and the parent populations, and to observe correlations between traits. Comparisons were done using ANOVA and linear regression. Statistics showed that all progeny populations had lower fiber yield than elite parents. This is expected and shows that alleles for traits important in the improvement of cotton are segregating and can be mapped. We also found that P017 had significantly longer fibers than its elite parent, which shows that it is possible to introgress beneficial alleles into elite cultivars from exotic sources. Acrylamide gels are currently being analyzed for differences in short sequence repeat (SSR) markers between populations. In addition, genotyping by sequencing is ongoing to look for SNPs associated with these phenotypes.

PO0601: Cotton

Intergenomic Gene Transfer in Diploid and Allopolyploid *Gossypium*

Nan Zhao¹, Corrinne E. Grover², Zhiwen Chen¹, Jonathan F. Wendel² and **Jinping Hua**^{1,3}, (1)China Agricultural University, Beijing, China, (2)Iowa State University, Ames, IA, (3)China Agricultural University, BEIJING 100193, China

Background: Intergenomic gene transfer (IGT) between nuclear and organellar genomes is a common phenomenon during plant evolution. *Gossypium* is a useful model to evaluate the genomic consequences of IGT for both diploid and polyploid species. Here, we explore IGT among nuclear, mitochondrial, and plastid genomes of four cotton species, including two allopolyploids and their model diploid progenitors (genome donors, *G. arboreum*: A₂ and *G. raimondii*: D₅).

Results: Extensive IGT events exist for both diploid and allotetraploid cotton (*Gossypium*) species, with the nuclear genome being the predominant recipient of transferred DNA followed by the mitochondrial genome. The nuclear genome has integrated 100 times more foreign sequences than the mitochondrial genome has in total length. In the nucleus, the integrated length of chloroplast DNA (cpDNA) was between 1.87 times (in diploids) to nearly four times (in allopolyploids) greater than that of mitochondrial DNA (mtDNA). In the mitochondrion, the length of nuclear DNA (nuDNA) was typically three times that of cpDNA. *Gossypium* mitochondrial genomes integrated three nuclear retrotransposons and eight chloroplast tRNA genes, and incorporated chloroplast DNA prior to divergence between the diploids and allopolyploid formation. For mitochondrial chloroplast-tRNA genes, there were 2-6 bp conserved microhomologies flanking their insertion sites across distantly related genera, which increased to 10 bp microhomologies for the four cotton species studied. For organellar DNA sequences, there are source hotspots, e.g., the *atp6-trnW* intergenic region in the mitochondrion and the inverted repeat region in the chloroplast. Organellar DNAs in the nucleus were rarely expressed, and at low levels. Surprisingly, there was asymmetry in the survivorship of ancestral insertions following allopolyploidy, with most *numts* (nuclear mitochondrial insertions) decaying or being lost whereas most *nupts* (nuclear plastidial insertions) were retained.

Conclusions: This study characterized and compared intracellular transfer among nuclear and organellar genomes within two cultivated allopolyploids and their ancestral diploid cotton species. A striking asymmetry in the fate of IGTs in allopolyploid cotton was discovered, with *numts* being preferentially lost relative to *nupts*. Our results connect intergenomic gene transfer with allotetraploidy and provide new insight into intracellular genome evolution.

Keywords: Intergenomic gene transfer, Allopolyploidization, *Gossypium*, Mitochondrial genome, Chloroplast genome, *NuMt*, *NuPt*

Submitter's E-mail Address: jinping_hua@cau.edu.cn

PE0602: Cotton

Transposable Element Dynamics in the Subgenomes of Allotetraploid Cultivated Cotton (*G. hirsutum* L.)

Bryan Naidenov¹, Charles Chen¹, Kevin R. Cushman², Alexander Lim¹, Akash Gupta³, Lori L. Hinze⁴, Junghyun Shim⁵, Megan Sweeney⁶, Paul Chomet⁷ and Benildo G. de los Reyes², (1)Oklahoma State University, Stillwater, OK, (2)Department of Plant and Soil Science, Texas Tech University, Lubbock, TX, (3)California State University, Northridge, CA, (4)USDA-ARS, Southern Plains Agricultural Research Center, College Station, TX, (5)Plant & Soil Science, Texas Tech University, Lubbock, TX, (6)BASF Corporation, Tech 3, Morrisville, NC, (7)NRGene Ltd, Ag Strategy, St. Louis, MO

The American cultivated cotton (*Gossypium hirsutum* L.) is a product of long evolutionary history initiated by the allotetraploidization of the ancient diploid AA and DD genome progenitors, followed by selection for fiber-elongation potentials during domestication. As comparative genome sequencing continues to unravel the intricacies of genome evolution in the genus, it becomes critical to address the question regarding the role of transposable elements (TE) in the structural and functional changes in each subgenome after allotetraploidization and domestication. Through an extensive *de novo* TE annotation of the newly assembled RefSeq genome for the stress-tolerant GDRS reference accession SA1766, we uncovered substantial evidence of differential genome rearrangements through the loss and gain of Long Terminal Repeat (LTR) families of TE. As an example, we found about 88% reduction in copy number of the LTR2778-subfamily in the A-subgenome of SA1766. The most significant reduction was established for chromosome-A03, which also represents the most substantial genome downsizing (19.6% reduction) compared to *G. arboreum*. On the other hand, LTR2778-subfamily appeared to have been recently acquired in the D-subgenome of SA1766 with enrichment on chromosome-D08, which is also among the chromosomes with the most significant expansion. Given the highest rejected substitution rate on chromosome-A03, we hypothesize that the removal of LTR families, possibly through several mechanisms including illegitimate recombination, may have been a critical event during domestication that provided a means for rapidly purging potentially deleterious mutations in the genic space.

PO0603: Cotton

Extended Haplotypes Are Associated with Population Differentiation and Environmental Adaptation in Cotton (*Gossypium hirsutum*)

Panhong Dai^{1,2}, **Shoupu He**¹, Gaofer Sun³ and Xiongming Du¹, (1)State Key Laboratory of Cotton Biology, Institute of Cotton Research of the Chinese Academy of Agricultural Sciences, Anyang, China, (2)Agricultural College, Yangtze University, Jingzhou, China, (3)Department of Computer Science and Information Engineering/Data Mining Institute, Anyang Institute of Technology, Anyang, China

Population divergence induced by large-scale chromosome structural variations (SVs) is suggested the major force for new species evolution and ecological adaptation in various animals and plants. In this study, we evaluated the genetic basis of environmental adaptation by integrating the genotype of 419 diverse *Gossypium hirsutum* accessions and their long-term environmental associated variables. Our results demonstrated that the population differentiation in *G. hirsutum* was mainly caused by genomic divergence on chromosome A06 and A08. Three extended haplotypes in the divergent regions were significantly associated with environmental adaptation. Our large-scale population genetic analysis firstly revealed the cause of population divergence in *G. hirsutum*, as well as its consequences of multi-phenotypic variations to adapt local environments. This finding provides new insight into the genetic basis of environmental adaptation in cotton, which could accelerate the developing of targeted cotton genotypes to face climate change in future cotton breeding.

PE0604: Cotton

Development and Characterization of A2D1 Chromosome Segment Substitution Lines in Cotton

Jamshaid A. Junaid¹, Ammani N. Kyanam¹, Selfinaz K. Veliloglu¹, Robert Vaughn², Wayne Raska³ and David M. Stelly⁴, (1)Department of Soil & Crop Sciences, Texas A&M University, College Station, TX, (2)Texas A&M University, Texas A&M Agrilife Research, College Station, TX, (3)USDA Forest Service, Southern Research Station, College Station, TX, (4)Department of Soil & Crop Sciences, Texas A&M University, Texas A&M Agrilife Research, College Station, TX

Cotton is a major agricultural commodity and is produced worldwide in tropical, subtropical, and temperate latitudes. High-yielding cultivated Upland cotton cultivars have low genetic variation due to natural and human influences, including its evolutionary origin 1-2 MYA as a new polyploid, recent domestication, adaptation to new production areas through selection, and breeding-based artificial selection. The extremely low genetic diversity found among Upland cottons constrains opportunities for improvements by breeding based on conventional elite-by-elite crosses. The availability of A2D1 synthetic tetraploid that involves the germplasm of diploid species *G. arboreum* and *G. thurberi* provides an opportunity to widen the genetic base of upland cotton A and D sub-genomes, and potentially to improve diverse traits through introgression with the upland cotton. This introgression of a synthetic tetraploid germplasm can significantly increase the breadth of genetic variation available among Upland cottons. However, it is practically impossible to predict whether a specific mutation would be helpful, neutral or deleterious in the genetic background of upland cotton. To effectively harness the genome of synthetic tetraploid, one method is the development of Chromosome Segment Substitution Lines (CSSLs) by modified backcross-inbreeding. Each CSSL would be selected by markers to contain one to several sub-chromosomal introgressed alien segments in an isogenic background of *G. hirsutum*(TM-1). Groups of CSSLs would be chosen on the basis of markers such that they jointly "capture" all or most of an alien genome. We report here progress towards a *G. hirsutum* (TM-1) groups of CSSLs containing germplasm from synthetic tetraploid (A2D1). These CSSLs will provide a mighty tool for introgression and characterization of synthetic tetraploid (A2D1) germplasm. The CSSLs (BC5Sn lines) will be high-density genotyped, cytogenetically characterizes meiotic behavior and tested for fiber (HVI, APHIS) and other traits.

PO0605: Cotton

Genome-Wide PCR-Based Simplex Assays for Genetic Diversification of Upland Cotton

Selfinaz K. Veliloglu¹, Luis M. De Santiago², Bree A. Vculek³, David B. O'Krafka³, Robert Vaughn⁴ and David M. Stelly⁵, (1)Department of Soil & Crop Sciences, Texas A&M University, College Station, TX, (2)Texas A&M University, College Station, TX, (3)TAMU, College Station, TX, (4)Texas A&M University, Texas A&M Agrilife Research, College Station, TX, (5)Department of Soil & Crop Sciences, Texas A&M University, Texas A&M Agrilife Research, College Station, TX

Upland cotton (*Gossypium hirsutum* L., $2n=52$) contributes significantly to the United States economy. Measurements have shown genetic diversity to be low among elite Upland cotton cultivars and breeding germplasm. Such low diversity can render an entire crop vulnerable to emerging biotic and abiotic threats and hinders improvement. Thus, improving genetic diversity by transferring novel alleles from wild species to cotton will likely enhance the success of downstream breeding efforts. To facilitate and expedite wide-cross breeding projects such as chromosome segment substitution line development, we have developed a genome-wide panel of map-spaced (cM) simplex SNP marker assays. Using ~18,000 mapped interspecific SNPs from the CottonSNP63K array and new AD genome sequence assemblies, we developed 10 or more evenly spaced simplex assays per linkage group (26). About 500 SNP assays, either "Kompetitive Allele Specific PCR" (KASP) or "PCR Allelic Competitive Extension" (PACE), were chosen according to their map locations and their abilities in plate-based assays to distinguish upland cotton, *Gossypium hirsutum* L., from three germplasm donor species, *Gossypium mustelinum*, *Gossypium tomentosum* and *Gossypium barbadense*, plus their corresponding F₁s, indicating likely utility for backcross introgression. This set of assays provides genome coverage at intervals of essentially 15-cM or less, i.e., sufficient for many interspecific breeding needs. Evaluations with 96.96 Dynamic Arrays™ and the Fluidigm EP1™ platform for relative positioning of genotype-specific clusters and cluster tightness, based on analysis of ~90 interspecific F₂ genotypes (~30 F₂ / donor species) indicated about 70% of the above assays will likely enable accurate genotyping of self-generations. Support from Cotton Inc. 13-466TX, 18-201, NSF IOS1739092 and the Republic of Turkey Ministry of National Education is gratefully acknowledged.

PE0606: Cotton

Development and Characterization of a Reciprocal Set of Interspecific Near Isogenic Lines in Cotton

Jeevan Adhikari¹, Rahul Chandnani², Ellen Regina E Skelton³ and Andrew H. Paterson³, (1)The University of Georgia, Athens, GA, (2)University Of Georgia, Athens, GA, (3)Plant Genome Mapping Laboratory, University of Georgia, Athens, GA

Cotton (*Gossypium spp.*) is a major source of natural fiber for textile industries around the globe. Mature cotton fibers, produced from tiny seed trichomes during four overlapping stages of development (fiber initiation, cell elongation, secondary cell wall deposition and maturation), are valued for their quality as defined by length, strength, fineness, elongation and uniformity. A lot of physiological changes, driven by associated transcriptome alterations, occur during these four stages of growth and development of cotton fiber. Identification and understanding of these alterations are important to dissect the stages involved in transforming primitive trichomes to the economically important fibers of modern cotton cultivars. To better understand the stages involved in transformation of the epidermal cell into mature cotton fiber and identify associated transcriptomic alterations, we are constructing a reciprocal set of near-isogenic lines (NILs) using Acala Maxxa (*G. hirsutum*) and Pima S6 (*G. barbadense*) as the parents, each of which contains one and only one introgressed segment from the donor genotype, but collectively covering the entire donor genome. These NILs, each consisting about 0.5% of the donor genome in a reference background, also provide a powerful tool for genetically dissecting complex traits, increasing the precision with which phenotypic changes can be mapped to transcriptomic and genetic alterations. Genotyping by sequencing of BC₃F₁ lines revealed an average of 3.6 introgressed segments in the population with Acala Maxxa as the recurrent parent. Genotyping is ongoing for the other set of NILs with Pima S6 as the recurrent parent. Introgressions will be verified in the BC₃F₂ progenies of each line with SSR markers selected from the introgressed regions and verified NILs with single introgressions will be phenotyped and characterized for various fiber quality related traits.

PO0607: Cotton

QTL and Genetic Analysis Controlling Fiber Quality Traits Using Paternal Backcross Population in Upland Cotton

Lingling Ma and **Jinping Hua**, China Agricultural University, Beijing, China

F₁₄ recombinant inbred line (RIL) population was backcrossed to paternal parent for a paternal backcross (BC/P) population, deriving from one Upland cotton hybrid. Three repetitive BC/P field trials and one BC/M field trial were performed including both two BC populations and the original RIL population. In total, 24 novel QTLs are detected for fiber quality traits and among which 13 QTLs validated previous results. 35 quantitative trait loci (QTL) in BC/P populations explain 5.01% - 22.09% of phenotype variation (PV). Among the 35 QTLs, 23 QTLs are detected in BC/P population alone. Present study provides novel alleles of male parent for fiber quality traits with positive genetic effects. Particularly, *qFS-Chr3-1* explains 22.09% of PV in BC/P population, which increased 0.48 cN/tex for fiber strength. A total of seven, two, eight, two and six QTLs explain over 10.00% of PV for fiber length, fiber uniformity, fiber strength, fiber elongation and fiber micronaire, respectively. In RIL population, six common QTLs are detected in more than one environment: *qFL-Chr1-2*, *qFS-Chr5-1*, *qFS-Chr9-1*, *qFS-Chr21-1*, *qFM-Chr9-1* and *qFM-Chr9-2*. Two common QTLs of *qFE-Chr2-2* (TMB2386-SWU12343) and *qFM-Chr9-1* (NAU2873-CGR6771) explain 22.42% and 21.91% of PV. The region between NAU4034 and TMB1296 harbor 30 genes (379 kb) in A05 and 42 genes (49 kb) in D05 for fiber length along the QTL *qFL-Chr5-1* in BC/P population, respectively. In addition, a total of 142 and 46 epistatic QTLs and QTL × environments (E-QTLs and QQEs) were identified in RIL-P and BC/P populations, respectively.

Keywords: FIBER QUALITY TRAITS, COMMON QTL, PATERNAL BACKCROSS POPULATION, UPLAND COTTON

PE0608: Cotton

Genome-Wide Dissection of the General Combining Ability of Fiber Quality- and Yield-Related Traits in Upland Cotton

Xiaoli Geng, Institute of Cotton Research, Chinese Academy of Agricultural Sciences, Anyang, Henan, China

An evaluation of the general combining ability (GCA) may facilitate the selection of suitable parents for hybrid cotton breeding, but the genetic basis of the GCA has not been fully characterized by molecular techniques. In this study, 282 female parents were crossed with four male parents in accordance with the North Carolina II mating scheme to generate 1,128 hybrids. The parental lines were genotyped based on restriction site-associated DNA sequencing, and 97,419 filtered SNPs were used for a genome-wide association analysis involving the phenotype and GCA values of 8 fiber quality- and yield-related traits. The main results were as follows: (1) 19 accessions with a top 5% GCA value for more than one trait were identified as elite parents for hybrid cotton breeding; (2) 665 significant SNPs for GCA were identified, and 103 common SNPs, which were detected across at least two environments were located in 43 quantitative trait loci; (3) 26 QTLs were both detected for the phenotype and the GCA of the female parents. Overall, our results suggest that pyramiding the favorable loci for the GCA may improve the efficiency of hybrid cotton breeding.

PO0609: Cotton

Large Insertion-Deletion Fragment Activates a New Fuzzless Gene in *G. arboreum*

Xiongming Du, Institute of Cotton Research, Chinese Academy of Agricultural Science, Anyang, Henan, China

Cotton seeds are tightly surrounded by very short fuzz fiber. Fuzzless phenotype is beneficial for seed coating and germination. Here, by performing the genome-wide association study strategy, we isolated a ~6kb length inserted fragment (larINDEL_{Fuzzless}) located at the end of chromosome 8, composed by a ~4.6 kb length repeats aside and a ~1.2 kb length homologous fragment derived from chromosome 12, conferring both seed fuzzless phenotype and leaf hairless in *G. arboreum*. This distant insertion as an enhancer located at ~17 kb upstream of dominant gene *GaFuzzless* (*Ga08G0121*). Overexpression of *GaFuzzless* suppressed the trichome number in *Arabidopsis*. The ectopic expression and CRISPR/Cas9-mediated gene knock-out analyses suggested that *GaFuzzless* also negatively modulated the fuzz and leaf hair development in cotton. Expression and interaction analysis demonstrated that *GaFuzzless* was independent of R2R3-MYB/bHLH/WD40 trichome regulating system, and repressed the expression of fatty acid metabolism related genes involved in trichome/fiber development. Our findings identify a new regulator in fiber/trichome development and provide new insight into the importance of non-coding sequences in cotton.

PE0610: Cotton

Silencing of a LIM Gene By Cotton-Mediated RNAi Exhibit Enhanced Resistance Against *Apolygus lucorum*

Sijia Liang, Zhongping Xu and Shuangxia Jin, National Key Laboratory of Crop Genetic Improvement, Huazhong Agricultural University, Wuhan, Hubei, China

Cotton production is challenged worldwide by a diversity of pests, causing severe yield loss and economic damage on cotton production. In recent years, plant bugs (Miridae species) has developed into a major agricultural pest in BT cotton fields. But so far there is no effective strategy to against bugs. In this study, we report the identification of a *LIM* gene from *Apolygus lucorum* (*AILIM*), down-regulation of *AILIM* expression by injection of double-stranded RNA leads to muscle structural disorders, resulting in metamorphosis deficiencies and increased mortality. Plant-mediated RNA interference (RNAi) is emerging as an eco-friendly, efficient and reliable alternative strategy for pest management. Then we constructed a plant expression vector and obtained transgenic cotton capable of highly and stably expressed dsRNA of *AILIM* (*dsAILIM*), by *Agrobacterium*-mediated genetic transformation. In the field bioassay, *dsAILIM* transgenic cotton are high-efficiently protected from *A. lucorum* damage, with no detectable yield penalty. Therefore our study successfully provided an additional option to high-efficiently protect plants from damage caused *A. lucorum*. Beyond that, we created transgenic cotton pyramids that combine protection from RNAi

and Bt toxin, we hope that, the pyramid can be as a promising germplasm resource to resist various of cotton pests and lay the foundation for the development of the next generation of GM crops with resistance to herbivorous pests.

PO0611: Cotton

Molecular, Genomic and Cellular Characterization of Emerging Fusarium Wilt of Cotton Race 4 in Texas

Catherine Danmaigona Clement¹, Zunyong Liu¹, Liang Kong¹, Lin Zhang¹, Steve Hague¹, Thomas Isakeit¹, Ping He¹, Jane Dever² and Libo Shan¹, (1)Texas A&M University, College Station, TX, (2)Texas A&M AgriLife Research and Extension Center, Lubbock, TX

Fusarium wilt disease of cotton race 4 (*Fov4*), poses a strategic threat to U.S upland cotton production by the recent expansion of *Fov4* from California into Texas in 2017. *Fov4* is difficult to control as the hyphae resides in the woody vascular tissues protected from fungicides with chlamydospores that can survive in soils forever. Resistance in upland cotton has not been identified and commercial varieties are not available. There is a critical need to understand the disease mechanism and develop upland cotton with resistance to *Fov4*. On the pathogen side, we need to understand the extent of the genetic diversity and parthenogenesis of different *Fov* isolates in the field and compare with existing *Fov* isolates. For this, we morphologically and molecularly characterized Texas isolates from 2017 and 2018 and performed whole-genome sequencing of 30 *Fov* isolates using MinION and illumina sequencing. Our study revealed unexpected genetic and virulence diversity of Texas *Fov* isolates. For example, some Texas *Fov4* isolates are more virulent than classical California race 4 CA14, whereas some isolates are non-pathogenic. To facilitate infection studies, we established a fast, quantitative and highly reproducible seed soak disease assay to determine cotton response to different isolates of *Fov4*. To reveal the infection process at the cellular level, we tagged several *Fov* isolates with green fluorescent protein (GFP) to monitor the fungal attachment, penetration, colonization and establishment by live-cell imaging. Our study reveals that *Fov4* mainly colonizes the root surface of the tolerant pima cotton with a much-delayed multiplication rate while it rapidly multiplies in the vascular tissues on the susceptible cotton cultivars causing seedling death.

PE0612: Cotton

A Novel Variant of Gh_D02G0276 Is Required for Root-Knot Nematode Resistance on Chromosome 14 (D02) in Upland Cotton

Martin J. Wubben¹, Gregory Thyssen², David D. Fang³, Dewayne Deng⁴, Jack McCarty⁵, Ping Li⁶, Md Sariful Islam⁷ and Johnie N Jenkins¹, (1)USDA-ARS, Mississippi State, MS, (2)Cotton Fiber Bioscience & Cotton Chemistry and Utilization Units, New Orleans, LA, (3)Cotton Fiber Biosciences Research Unit, USDA-ARS-SRRC, New Orleans, LA, (4)USDA ARS, Mississippi State, MS, (5)Genetics & Sustainable Agriculture Research Unit, USDA-ARS, Mississippi State, MS, (6)Cotton Fiber Bioscience Research Unit, USDA-ARS-SRRC, New Orleans, LA, (7)Sugarcane Production Research Unit, USDA-ARS, Canal Point, FL

The southern root-knot nematode [RKN; *Meloidogyne incognita* (Kofoid & White)] remains the primary yield-limiting biotic stress to Upland cotton (*Gossypium hirsutum* L.) throughout the southeastern United States. While useful genetic markers have been developed for two major RKN resistance loci on chromosomes 11 (A11) and 14 (D02), these markers are not completely effective because the causative genes have not been identified. Here, we sequenced 550 recombinant inbred lines (RILs) from a multi-parent advanced generation intercross (MAGIC) population to identify five RILs that had informative recombinations near the D02-RKN resistance locus. The RKN resistance phenotypes of these five RILs narrowed the D02-RKN locus to a 30-kb region with four candidate genes. We conducted virus induced gene silencing (VIGS) on each of these genes and found that Gh_D02G0276 was required for suppression of RKN egg production phenotype conferred by the Chr. D02 resistance gene. The resistant lines all possessed an allele of Gh_D02G0276 that showed non-synonymous mutations and was prematurely truncated. Furthermore, a Gh_D02G0276-specific marker for the resistance allele variant was able to identify RKN resistant germplasm from a collection of 367 cotton accessions. The Gh_D02G0276 peptide shares similarity with domesticated *hAT*-like transposases with additional novel N- and C-terminal domains that resemble the target of known RKN effector molecules and a prokaryotic motif, respectively. The truncation in the resistance allele results in a loss of a plant nuclear gene specific C-terminal motif, potentially rendering this domain antigenic due to its high

homology with bacterial proteins. The conclusive identification of this RKN resistance gene opens new avenues for understanding plant resistance mechanisms to RKN as well as opportunities to develop more efficient marker assisted selection in cotton breeding programs.

PO0613: Cotton

Differentially Expressed Genes (DEGs) in Response to Reniform Nematode in Upland Cotton, *Gossypium hirsutum*

Shalini Priya Etukuri¹, Abreeotta Williams², Zachary B. Gossett¹, Varsha C. Anche¹, Govind C. Sharma¹ and Venkateswara R. Sripathi¹, (1)Alabama A&M University, Normal, AL, (2)ALABAMA A&M UNIVERSITY, NORMAL, AL

Rotylenchulus reniformis, commonly known as reniform nematode (RN), is an agriculturally devastating parasitic nematode, which has drawn the world's attention. It is a semi-endoparasitic nematode (partially inside the root) that is a major yield-limiting pest of multiple crops in the tropics and sub-tropics, including upland cotton (*Gossypium hirsutum*). Reniform nematode (RN) is a menace to global food and fiber production, especially in combination with associated fungal and bacterial diseases. Throughout the world, cotton is the most essential natural textile fiber and a prominent oilseed crop. In the USA, the annual loss of cotton due to RN is approximately 525,000 bales in a total of \$130 M. Our research objectives are: generating resources to limit RN infestation; confirming the accuracy and field to field reproducibility of rapid screening procedures for detection and quantification of RN; and analyzing the defense response genes in RN susceptible and resistant to *G. hirsutum* genotypes. Two genotypes of *G. hirsutum* (Lonren-1 and FiberMax) were germinated in 40% sand and 60% clay. Twenty days after germination, the plants were exposed to different concentrations of RN (0, 5k, and 50k nematodes/quart). Twenty-one days after infection, roots were collected and stained with acid fuchsin for confirming the RN infection. Then, RNA was isolated, quality-tested, and used for sequencing and bioinformatics analyses. The transcriptome analysis in Loren-1 and FiberMax identified over twenty thousand differentially expressed genes (DEGs), when compared to the control plants. Among these, two genes with a prominent role in biotic stresses and 17 genes associated with resistance were identified. Additional work is required to validate the predicted genes related to RN resistance.

PE0614: Cotton

Physiological Networks Rather Than Cellular Na⁺ Transport and Sequestration Explain Phenotypic Variation for Salt Tolerance Potential Across Cultivated *Gossypium* Germplasm

Kevin R. Cushman¹, Isaiah Pabuayon¹, Lori L. Hinze², Megan Sweeney³ and Benildo G. de los Reyes¹, (1)Department of Plant and Soil Science, Texas Tech University, Lubbock, TX, (2)USDA-ARS, Southern Plains Agricultural Research Center, College Station, TX, (3)BASF Corporation, Tech 3., Morrisville, NC

After the osmotic shock from exposure to excess salt, toxic ions accumulate in tissues of salt stressed plants. In glycophytes, the ability to mobilize or sequester toxic ions has been associated with tolerance. Mobilization and sequestration of excess Na⁺ involves three transport mechanisms namely SOS1 (plasma membrane H⁺/Na⁺ antiporter), NHX1 (vacuolar H⁺/Na⁺ antiporter), and HKT1 (Na⁺/K⁺ transporter in vascular tissue). Plants adapted to salty environments are assumed to have a better capacity to maintain cytoplasmic ionic balance.

As a moderately salt tolerant salt tolerant crop plant, cotton (*Gossypium hirsutum* L.) has shown variation in response to salinity stress. Phenotypic variation to salt was investigated across a comparative panel representing the spectrum of genetic diversity based on microsatellite marker data in improved cotton germplasm in relation to the spatio-temporal patterns of Na⁺ accumulation. The goal was to discover physio-morphometric attributes and their interactions that may contribute to overall tolerance capacity in context of the contributions of *GhHKT1*, *GhSOS1*, and *GhNHX1* to Na⁺ homeostasis. Multi-dimensional physio-morphometric attributes were investigated in a univariate and multivariate statistical context, as well as, the relationship between variables using structural equation modeling. Results showed that mobilized or sequestered Na⁺ contribute to the baseline salinity tolerance, but the observed variance in overall stress tolerance potential across the comparative panel were more significantly associated to antioxidant capacity, maintenance of stomatal conductance, chlorophyll concentration, Ca²⁺ and Mg²⁺ content, and other physiological interactions.

PO0615: Cotton

Cotton DNA Methylation and Its Analysis Under the Salt- & Draught-Stress

Wuwei Ye, Institute of Cotton Research, CAAS, China, Anyang, China

DNA methylation, an important component of epigenetics induced usually by adversity, plays a vital role in the response to various stresses including drought and salt. A methylation-sensitive amplification polymorphism method based on capillary electrophoresis was used to explore the epigenetic mechanisms of salt tolerance and heterosis in Upland cotton (*Gossypium hirsutum* L.), and the results indicated that hypermethylation and demethylation could be an important mechanism to resist the stresses. And the demethylation could be the mechanism to explain heterosis in cotton hybrid. The results of whole genome methylation sequencing showed high DNA methylation density usually occurs in promoter regions and transposons areas. Methylated cytosines in different sequence contexts (CG, CHG and CHH) have different functions and methylation levels. And the results also showed methylated cytosines in asymmetric CHH sequence context are dynamic, being mostly related to stresses. Combined with transcriptome data, we found long non-coding RNAs (lncRNAs) may involve in the regulation of DNA methylation in response to drought stress. All these results could provide theoretical reference value for the mechanism research of tolerance in cotton.

PE0616: Forest Trees

TreeGenes and CartograTree: Tools for Forest Tree Genomics

Emily Grau¹, Sean Buehler², Nic Herndon², Peter Richter², Risharde Ramnath³, Margaret Staton⁴, Stephen P. Ficklin⁵ and Jill L. Wegrzyn², (1)Department of Plant Sciences, Davis, CA, (2)Department of Ecology and Evolutionary Biology, University of Connecticut, Storrs, CT, (3)University of Connecticut, Storrs, CT, (4)University of Tennessee, Knoxville, Knoxville, TN, (5)Dept of Horticulture, Washington State University, Pullman, WA

TreeGenes is designed to serve the specialized needs of the forest tree research community by pairing data resources, for both model and non-model species, with customized analysis tools. TreeGenes is built on open source tools, including Tripal, GMOD applications, and Galaxy to maximize efficiency, functionality, and interoperability.

Genetic/genomic, phenotypic, and environmental data is sourced from primary repositories as well as direct user submissions using the Tripal Plant PopGen Submission (TPPS) module, which allows users to submit population genetic data and metadata. Over 1900 species from 157 genera are represented in the database.. Additional data available includes literature, colleague listings, and community resources.

TreeGenes hosts tools adapted to the challenges posed by the complex genetics of forest trees. High performance computational resources to perform the analyses are hosted by TreeGenes through the Tripal Galaxy module, allowing users anywhere to run the common bioinformatic applications. Users can perform rapid sequence similarity search using the TSeq tool, visualize orthologous gene sets in a phylogenetic context with the OrthoQuery, and perform landscape and association mapping analysis within CartograTree.

CartograTree, a map-based framework, provides an efficient interface for researchers to analyze georeferenced tree populations in conjunction with their associated genotypic and phenotypic metrics. Environmental layers such as WorldClim and soil types are integrated with data from TreeGenes as well as external databases like Dryad and TreeSnap. CartograTree is connected to Galaxy to allow users to run open-source tools such as Sambada, Plink, and Structure.

PO0617: Forest Trees

The Hardwood Genomics Project: An Online Database for Tree Genetic and Genomic Data

Matthew Huff, University of Tennessee, Knoxville, TN

As genetic and genomic data for woody tree species increases, the need for this data to be easily accessible for fellow researchers also increases. At present, much of this data is only available in raw or unannotated formats, if it

is available at all. The Hardwood Genomics Website (HWG) is dedicated to addressing the growing need for a central database for woody tree genomes. Our site houses genomes and transcriptomes of trees unavailable on any other platform and also provides searchable functional annotation for genes and transcripts. To further characterize gene sequences, we house gene expression data from high throughput RNASeq experiments, allowing users to determine how their genes are affected by environmental conditions or if their genes are differentially expressed between tissues. We have 10 species with expression datasets currently represented and biocurators actively adding more. The site allows users to see genes that were statistically significant in the expression experiment and to generate heatmaps for genes of interest. We have also integrated a number of tools for researchers to easily access our data, including a powerful search engine, BLAST sequence similarity searching, and JBrowse for genome browsing. This system allows users to use data available on HWG, as well as data uploaded from their computer, as input in a workflow. Following the successful integration of high-throughput transcriptome data, we continue to welcome new data submissions, suggestions, and partnerships to continue development. HWG is supported by NSF Awards #1444573.

PE0618: Forest Trees

Evo-Devo of Secondary Growth Traits in the Seed Plant Lineage

Danielle Roodt, University of Pretoria, Pretoria, South Africa

Wood (secondary xylem from vascular cambium) is found in the gymnosperm and angiosperm lineages but has been lost in the monocotyledons. Although wood formation is ancestral to the seed plants and was likely lost at the base of the monocots, differences between the eudicots and most gymnosperm lineages are apparent at the anatomical, cell wall and molecular levels. One gymnosperm lineage, the Gnetales, shares similarities with the eudicots but not other gymnosperms. We aimed to gain insight into the cause of this gain and loss of woody characteristics in the Gnetales and monocots, respectively. We analysed orthologs of genes preferentially expressed in developing xylem of eudicot trees across the angiosperms, constructed a regulatory model of early vascular cambial cell identity and differentiation and investigated its conservation in the sequenced genomes of 26 eudicot and seven monocot species and the early diverging angiosperm *Amborella trichopoda*. We also constructed high quality gene catalogues for representative species from the three non-coniferous gymnosperm lineages from transcriptome data of xylogenetic and contrasting tissues, allowing for comparative transcriptomics and investigations of the gene networks underlying xylogenesis. Furthermore, we performed wood morphology and wood chemistry comparisons between the seed plant lineages and identified candidate genes for further elucidation of convergently evolved or ancestrally retained traits. These results broaden our understanding of the evolution of the highly divergent seed plant lineage and provide candidate genes that may have led to the loss of wood formation in the monocot lineage, and the development of “angiosperm-like” traits in the Gnetales.

PO0619: Forest Trees

The Open Chromatin Regulation of Complex Traits

Nathaniel W. Ellis, Annette M. Fahrenkrog, Kelly Balmant, Daniel Conde, Christopher Dervinis, Jerald D. Noble, Marcio F. R. Resende, Jason Smith, Brad Barbazuk and Matias Kirst, University of Florida, Gainesville, FL

Populus deltoides (Eastern cottonwood) is a short-rotation woody crop with strong potential for bioenergy production because of its fast growth and wide adaptation to the Midwest and Southern United States. *P. deltoides* is cultivated worldwide for lumber and biomass, but development of new cultivars is hindered by lengthy breeding cycles and difficulties in phenotyping traits such as disease resistance and yield. The long-term goal of this research is to uncover and apply genomic information to guide and accelerate improvement of poplar cultivars. In order to uncover genes regulating critical traits we are characterizing a genetically unrelated population of 425 *P. deltoides* individuals. Genome-wide association studies (GWAS) have been now expanded to include growth properties measured under field conditions and disease resistance to *Sphaerulina musiva* in this population. In addition, we are using the Assay for Transposase-Accessible Chromatin combined with next-generation DNA sequencing (ATAC-seq) to uncover open chromatin regions across the *P. deltoides* genome and variants contained within them. Our hypothesis is polymorphisms within these accessible regions of the *P. deltoides* genome contribute a larger proportion of the phenotypic variance. Evidence indicates that, contrary to results obtained in species with highly methylated genomes, regions of open chromatin are not enriched for loci that control complex traits in *Populus*.

PE0620: Forest Trees

Development of a 20K SNP Array and Evaluation for High-Density Mapping in *Pinus thunbergii*

Tomonori Hirao¹, Koji Matsunaga², Kentaro Mishima³ and Yuichiro Hiraoka³, (1)Forest Bio-Research Center, Forestry and Forest Products Research Institute, Hitachi, Ibaraki, Japan, (2)Kyushu Breeding Office for FTBC, Goshi, Japan, (3)Forest Tree Breeding Center, Forestry and Forest Products Research Institute, Hitachi, Ibaraki, Japan

Both a genetic resource and the development of genomic tools, such as reference genomes and molecular genetic markers, are important to enable faster progress in plant breeding. To promote resistance breeding to pine wilt disease in Japanese black pine, we designed a high-density, high-efficiency and robust single nucleotide polymorphism (SNP) array in *Pinus thunbergii*, with the main objectives of conducting evaluation of genetic diversity, linkage analysis, and genome-wide association studies. By applying polymorphism information detected from RNA sequencing in 8 varieties, we were able to mount the most robust and informative SNPs on the Applied Biosystems Axiom 20 K Genotyping Array, currently the densest SNP array in *Pinus thunbergii*. Preliminary evaluation of this 20 K array in 96 varieties from the FTBC-FFPRI, identified 7,111 SNPs as high quality and polymorphic (PolyHighResolution). We further used the Axiom 20 K Genotyping Array to construct high-density linkage maps in a bi-parental population, and to make a direct comparison with a linkage map in *Pinus thunbergii*, which suggested that the SNP array is a more informative than previously constructed a linkage map.

PO0621: Forest Trees

Functional Genomics of Induced Resin Duct Cells in *Pinus taeda*

Daniel D. Ence, Adam D Sims, Mallory Morgan and Gary F. Peter, University of Florida, Gainesville, FL

Terpene production and storage is an ancient defense mechanism in plants and is of commercial interest. Terpenes produced in loblolly pine are produced in specialized resin duct epithelial cells and stored in extracellular lumens of resin ducts. Exogenous application of methyl jasmonate has been shown to induce the resin defense response as well as new resin duct formation. Trees with increased terpene production and resin duct formation would be of value through increasing disease-resistance and through increasing production of valuable chemicals.

Resin duct epithelial cells arise as initial cells in the cambial zone. In response to methyl jasmonate treatment, initial cells begin to differentiate to resin duct epithelial cells as they move into the xylem. The determinant of cell fate for initial cells is currently unknown. Also unknown is how cell fate is related to the upstream methyl jasmonate signaling pathway and the downstream increase in terpene biosynthesis.

In this work we integrated RNAseq data from methyl jasmonate-treated pine seedlings with GWAS data from a pine breeding population that had been phenotyped for resin duct number, resin flow, and accumulation of specific terpenes. We identified differentially expressed transcripts from genes in the terpene biosynthesis and jasmonic acid signaling pathways. We also annotated SNPs in terpene biosynthesis genes that were associated with terpene phenotypes. Together this data adds to knowledge of which genes that are involved in methyl jasmonate-mediated terpene formation and terpene production phenotypes.

PE0622: Forest Trees

Identification of Candidate Major Resistance Genes in the *Pinus taeda* L. Genome

Daniel D. Ence¹, K. Smith², Matias Kirst¹, Leandro G. Neves³, Jill L. Wegrzyn⁴, C. Dana Nelson² and John M Davis¹, (1)University of Florida, Gainesville, FL, (2)Southern Institute of Forest Genetics, Saucier, MS, (3)RAPiD Genomics, Gainesville, FL, (4)Department of Ecology and Evolutionary Biology, University of Connecticut, Storrs, CT

Fusiform rust is a disease of southern pines (where it causes galls on stems and branches) and on oaks (where it causes minimal leaf damage). Fusiform rust is a major disease threat to the timber industry in the US where it causes yield losses that exceed US\$100M/year. A high priority for breeders and forest managers is to identify candidate

resistance genes in loblolly pine. However, identifying the specific loci that regulate phenotypic traits in conifers is a major undertaking because of their very large genomes. During the process of annotating the genome of *Pinus taeda*, an expressed sequence tag (EST) was identified that contains a single nucleotide polymorphism (SNP) mapping to the locus (*Fr1*) that interacts with the fungal avirulence gene, *Avr1*. This EST aligns to a transcript from RNA-sequencing data and a TIR-NB-LRR protein, thus identifying it as a candidate *Fr1* gene. Here we present the results of work mapping Fusiform rust resistance locus 1 (*Fr1*) in the *Pinus taeda* genome. We conducted bulk segregant analysis of next-generation sequence data in pine. Half-sibling progeny from a resistant mother were phenotyped as either resistant or susceptible to CQF. These progeny were sequenced with a custom sequence-capture method targeting a genomic region linked to resistance by prior work. By identifying candidates for a resistance gene in this pathosystem, we will discover markers that will guide breeding and deployment of resistant pine.

PO0623: Forest Trees

Long-Read Assembly of Norway Spruce and Scots Pine

Niklas Mähler¹, Tomas A Larsson², Manfred Grabherr², Bastian Schiffthaler¹, Nicolas Delhomme¹, Björn Nystedt² and Nathaniel R. Street¹, (1)Umeå Plant Science Centre, Umeå University, Umeå, Sweden, (2)SciLifeLab, Uppsala, Sweden

Conifers are among the oldest organisms on this planet. They are of great ecological importance as well as being economically important species. We are performing long-read sequencing of the 20 Gb Norway spruce genome (*Picea abies*) and the 24 Gb Scots pine genome (*Pinus sylvestris*) in order to enable acceleration of the breeding programs for these species. The genome of Norway spruce has been assembled using 30X coverage of PacBio reads with the wtdbg2 assembler. Our current draft assembly has an N50 of 150 Kb before scaffolding, which has substantially improved contiguity of the regulatory regions surrounding genes, as measured by mapping back 21,300 full-length transcripts obtained with PacBio Iso-Seq data from a mixed pool of tissues. We expect assembly contiguity to improve markedly after scaffolding and anchoring to our ultra-dense genetic map. The long term goals of the current project are to understand the genetic basis of variation in economic and adaptive traits by examining local adaptation in natural populations by GWAS, transplanting experiments and growth chamber experiments and to develop tools enabling genomic selection within breeding and deployment programs. This requires high quality, well annotated reference genomes coupled to targeted experiments to inform biological interpretation. We are also performing comparative studies with other conifer species and will utilise this to increase our understanding of how our two target species will respond to changes in abiotic stresses that are predicted to become more prevalent and extreme as a result of ongoing climate change.

PE0624: Forest Trees

The Genomes of Engelmann, Sitka, White Spruce and Their Ingress Provide Insights into Local Adaptation

Kristina Gagalova¹, Lauren Coombe¹, Diana Lin¹, Ka Ming Nip¹, Hamid Mohamadi², Justin Chu¹, S. Austin Hammond³, René L. Warren¹ and Inanc Birol¹, (1)Canada's Michael Smith Genome Sciences Centre, Vancouver, BC, Canada, (2)Weill Cornell Medicine, New York, NY, (3)University of Saskatchewan, Saskatoon, SK, Canada

Spruce (genus *Picea*) are coniferous evergreens primarily found in the Northern hemisphere. These trees dominate the natural landscape across Canada, representing an integral resource to the lumber industry. We have assembled and annotated the genomes of two important spruce species, Sitka (*P. sitchensis*; Q903 genotype) and Engelmann (*P. engelmannii*; Se404-851 genotype) and improved the genomes and annotations of white spruce (*P. glauca*; WS77111 genotype) and the British Columbian interior spruce (PG29 genotype). These four species represent a variety of phenotypes and range of growth characteristics.

Herein we compare the genome annotations of the four spruce species to related conifers and evaluate their degree of completeness using several metrics. We have used a list of high-quality gene annotations to investigate the species' phylogenetic relationships. We clustered the inferred protein sequences from annotations into orthogroups. Based on the reconstructed phylogeny, we have analyzed the changes in gene family sizes using maximum likelihood inference methods.

We observe that the Engelmann, white and interior spruce genomes have a large number of rapidly evolving genes compared to the other conifer genomes evaluated. Interestingly among the functional domains annotated, we have also identified several gene candidates relating to local adaptation, such as transcriptional factors, cytochrome P450, UDP-glucuronosyltransferases, heat-shock protein Hsp20 family, and ubiquitin gene family. Our analyses suggest that the gain/loss of particular gene families plays an important role in the local adaptation of these species.

PO0625: Forest Trees

Genome-Wide Association of Cold-Related Traits in Coastal Douglas-fir

Benjamin Wilhite¹, Daniela Puiu², Brad St Clair³, Marc Crepeau⁴, Steven L. Salzberg², Charles H. Langley⁴, Brian J Allen⁵, David B. Neale⁶ and Amanda de la Torre⁷, (1)Northern Arizona University, Flagstaff, AZ, (2)Johns Hopkins University, School of Medicine, Baltimore, MD, (3)USDA Forest Service Pacific Northwest Research Station, Corvallis, OR, (4)Department of Evolution and Ecology, University of California, Davis, Davis, CA, (5)Department of Plant Science, University of California, Davis, Davis, CA, (6)Department of Plant Sciences, University of California, Davis, Davis, CA, (7)Northern Arizona University, Flagstaff, AZ

This study aims to combine phenotypic, genotypic, and environmental data to gain insights into the genomic basis of cold adapted traits in coastal Douglas Fir (*Pseudotsuga menziesii* var. *menziesii*). Genome-wide SNPs obtained from whole-genome re-sequencing of Douglas fir individuals were used to test for associations between 22 phenotypic traits related to cold damage (e.g. electrolytic leakage), growth and resource partitioning (e.g. root to shoot ratio), and phenology (e.g. budburst). Individuals were also used to test for associations between 54 environmental variables (e.g. temperature and precipitation). Population structure analysis showed that two distinct groups exist within our study zone; one existing in Southern Oregon (type 2), with the other more dominant type throughout the rest of the sampled range, up to the Canadian border (type 1). The two types hybridize, and we found a small number of hybrids in Southern Oregon. The hybrids are all more closely related to type 1 than type 2, suggesting asymmetric gene flow from type 1 to type 2. The hybrids cluster separately from type 1 in ordination space, suggesting functional differences between the genetic clusters. Most of the significant phenotype associations (163 in total) were with traits related to growth and resource partitioning such as diameter, height after year two, and root length. We did not find any associations with cold related traits. Most of the GEA results (723 in total) were associated with July maximum temperature, sun exposure and distance to sea. There were no associations in common between the GWAS and GEA results.

PE0626: Forest Trees

Determining the Defensive Mechanisms in Green Ash (*Fraxinus pennsylvanica*) Resistant to Emerald Ash Borer (*Agrilus planipennis*)

Robert K. Stanley¹, David Carey², Jennifer Koch³, Margaret Staton⁴ and Jeanne Romero-Severson¹, (1)University of Notre Dame, Notre Dame, IN, (2)USDA Forest Service Northern Research Station, NRS-16, Delaware, OH, (3)USDA Northeastern Research Station, Delaware, OH, (4)University of Tennessee, Institute of Agriculture, Knoxville, TN

Emerald ash borer (EAB, *Agrilus planipennis*), an accidentally introduced Asian beetle, poses an acute threat to the native *Fraxinus* species in North America due to the trees' lack of functional resistance. EAB has killed hundreds of millions of green ash (*Fraxinus pennsylvanica*) and has spread through over half of the native range of green ash. A small number of green ash (<1%) termed "lingering" survive for years after all other local green ash have died. Our collaborators have used artificial infestations of grafted clones of these lingering trees to verify that most of these lingering phenotypes are due to reproducible quantitative resistance to EAB. They then generated full sib progeny of lingering x lingering and susceptible x susceptible parents. By using structured crosses, we can account for the confounding effects of genetic background and identify the potentially different resistance mechanisms seen in the phenotypes. These multiple resistance mechanisms can be 'stacked' or pyramided in a selective breeding program to produce trees with greater long-term resistance to EAB.

We employ a multi-faceted, interdisciplinary approach to examine the functional basis for resistance to EAB in lingering green ash. We use transcriptomics, proteomics and metabolomics to examine differences in gene

expression, proteins and secondary metabolites in susceptible green ash vs lingering green ash. We have applied this analysis to 175 progeny from lingering x lingering and susceptible x susceptible crosses in twelve different families. Our transcriptomics and metabolomics analyses of these structured populations reveals significant standing genetic variation within green ash population for metabolites. In addition we have conducted analyses that show 1) uninfested families with different parents have different metabolic profiles 2) the metabolic profiles of infested vs uninfested progeny within a single family are different and 3) within families, we have the potential to distinguish between the metabolic profiles of infested progeny with highest and lowest defensive responses. Additionally, discriminant analysis reveals that while there is an overall chemical response to infestation, this is distinguishable across families.

This first set of results allows us construct predictions of phenotypes based on metabolic profiles in a study of larger green ash structured populations. If our predictions are supported, we can design a test that will allow for a strong prediction of resistance from a small tissue sample. This will allow for higher throughput in collecting new potential lingering trees from the wild and testing progeny from our controlled crosses. By increasing the rate at which defensive traits in ash can be selected for in a targeted breeding program, we can produce green ash with enough resistance to restore green ash on the landscape and in our cities.

PO0627: Forest Trees

Investigating Genetic Signatures Associated with Reduced Mortality Against Emerald Ash Borer in Green Ash

Ava Fritz¹, Jeremy Bennett¹, Jeanne Romero-Severson², Megan Reid², Jennifer Koch³ and Jill L. Wegrzyn⁴, (1)University of Connecticut, Storrs, CT, (2)University of Notre Dame, Notre Dame, IN, (3)USDA Northeastern Research Station, Delaware, OH, (4)Department of Ecology and Evolutionary Biology, University of Connecticut, Storrs, CT

Emerald Ash Borer (*Agrilus planipennis*), an invasive pest that is threatening the sustainability of the entire North American ash tree population (*genus: Fraxinus*), has the ability to kill a susceptible tree within five years of detection. Efforts to prevent, treat, and remove afflicted trees are culminating to billions of dollars, and the eradication of this species is impacting urban and rural landscapes. Among impacted populations, individuals that survive longer have been identified as “lingering ash,” though these individuals account for less than 1% of the population. To complement efforts to examine metabolomic profiles of populations of lingering and susceptible trees, we conducted deep double-digest RAD-Seq across a phenotyped population of 100 green ash (*Fraxinus pennsylvanica*). The susceptible and lingering ash families were phenotyped in greenhouse studies with EAB larvae to calculate the percentage of larvae killed by host natural defense, as well as the mean weight of surviving larvae. Genome-wide association approaches were conducted with the high quality reference genome to examine loci or regions of the genome associated with partial resistance (or lower susceptibility) to EAB. Deep sequencing provides an opportunity to assess variants and compare with recently identified associations in critical enzymes involved in the phenylpropanoid pathway and chemical defense. The preliminary genomic information gathered here will be used in conjunction with existing efforts to intentionally breed green ash that is highly resistant to EAB and can be used to repopulate highly impacted areas.

PE0628: Forest Trees

Alternative Splicing Plays a Critical Role in Plant–Pathogen Interactions in *Populus*

Jin Zhang, Meng Xie, Zhenzhen Qiao, Jay Chen and Wellington Muchero, Oak Ridge National Laboratory, Oak Ridge, TN

Alternative splicing (AS) is the process through different splice sites in precursor messenger RNA to generate multiple mRNA isoforms. It plays a pivotal role in the flow of genetic information from DNA to proteins by expanding the coding capacity of genomes. Regulation of AS is as important as regulation of transcription to determine cell- and tissue-specific features, normal cell functioning, and responses of Eukaryota cells to external cues. However, the extent of genome-wide AS changes in population-wide is largely unknown in plants, especially in woody species. On the basis of a time-course RNA-Seq data of two *P. trichocarpa* genotypes (resistant and susceptible) infected by *Sphaerulina musiva* (an invasive fungal pathogen), we compared the differentially AS

events in the two genotypes and the time-course changes in each genotype. Through an integrated co-expression analysis using AS splicing level and gene expression level, we identified an intron retention event in a novel transcription factor is associated with the defense response in *Populus*. The transactivation assay indicated that this transcription factor is a transcription repressor. We tested five potential downstream genes (*WRKY70*, *WRKY43*, *HsfB3*, *S-LPK* and *PR5K*) and found that the intron retention in this transcription factor affects repressor activity to the downstream genes. These results indicate that alternative splicing plays a critical role in defense response in *Populus*.

PO0629: Forest Trees

Genome Shattering Patterns Analysis in Poplar

Weier Guo¹, Isabelle M. Henry¹, Luca Comai² and Meric Lieberman³, (1)UC Davis, Davis, CA, (2)Plant Biology and Genome Center, UC Davis, Davis, CA, (3)University of California, Davis, Davis, CA

Chromoanagenesis is a genomic catastrophe during which a single chromosome or localized chromosomal region undergoes a large number of complex chromosomal rearrangements. It is correlated with cancer, but has also been detected in plants in Arabidopsis aneuploids. The origin and rearrangement mechanism of chromoanagenesis, as well as its contribution to genetic architecture and genome evolution are not fully understood. In a previous study on gene dosage variation in poplar, we established an interspecific F1 population from a cross between a *P. deltoides* female and gamma-irradiated pollen from a *P. nigra* male. Among this F1 population, several genotypes exhibited genome shattering patterns, which were consistent with chromoanagenesis. Here, we used genome sequencing to analyze the genome architecture of these shattered *Populus* lines. We reconstructed sequence junctions from these shattered chromosomes in silico, and identified various junction types and fragment orientations. Analysis of alleles frequencies using SNPs loci suggested that all of the duplicated or triplicated fragments originated from *P. nigra*, and might have resulted from the pollen irradiation treatment. With the integration of breakpoint junctions and the prediction of sample fragments rearrangement, we determined that fragment remodeling was intra-chromosomal. Our results indicated that chromoanagenesis may be more prevalent than expected, and can be triggered by chromosomal breaks, such as those resulting from gamma irradiation.

PE0630: Forest Trees

Insights & Engineering of Symbiotic Nitrogen Fixation in Poplar

Thomas Irving¹, Lucas Maia¹, Sara Knaack¹, Daniel Conde², Matias Kirst², Sushmita Roy¹ and Jean-Michel Ané¹, (1)University of Wisconsin-Madison, Madison, WI, (2)University of Florida, Gainesville, FL

Current knowledge on plant symbioses with nitrogen fixing bacteria supports a single origin for the evolution of root nodules. Poplar is a close relative to the plants of this 'nitrogen fixing clade', and represents an excellent model for gain-of-nitrogen-fixation experiments. We show that poplar retains most genes considered necessary for nodulation, and describe a limited change approach to generate nodules capable of housing nitrogen fixing bacteria in a non-native host. We are working to alter promoter function to co-ordinate expression of two conserved signalling pathways to achieve this, focusing on the NIN/NLP transcription factor family and the hormone cytokinin. Insights gained from the poplar model can be used spread this agronomically useful symbiosis to more distantly related crop plants, and help increase yields in regions unable to access large-scale fertilization.

PO0631: Forest Trees

The Early Bud-Break Regulon in Poplar

Abdul Azeez, School of Forest Resources and Environmental Science, MTU, Houghton, MI

Bud dormancy is an adaptive strategy in perennials from temperate and boreal climates to survive unfavorable conditions during winter months. Bud dormancy is developed in fall, released after prolonged cold during winter and followed by reactivation of growth (aka bud-break) in response to warm temperatures in spring. Molecular regulation of dormancy release and bud-break are largely unknown. We have identified through activation tagging early bud-break dominant (*ebbD*) poplar mutants and accordingly the underpinning genes are called EARLY BUD-

BREAK (EBB). Previously we have reported the isolation and characterization of the EBB1 gene from one of these mutants, encoding an AP2/ERF transcription factor. We have now found that EBB1 directly represses the poplar SVL (SHORT VEGETATIVE PHASE-LIKE), a MADS-box protein which was recently found to negatively regulate bud-break in *Populus*. We also report the identification and characterization of the *ebb3D* mutant. The corresponding gene EARLY BUD-BREAK 3 (EBB3), encodes another AP2/ERF domain-containing transcription factor. EBB3 overexpressing lines showed early bud-break whereas, RNAi plants showed significantly delayed bud-break as compared to wild type control. We show that EBB3 is downstream of EBB1 and SVL. EBB1 positively, while SVL negatively regulates *EBB3* expression. Further, we show that EBB3's effect on bud-break is mediated by the regulation of the cell cycle. EBB3 directly and positively regulates the *CYCD3.1* gene, encoding an important checkpoint in the progression of the cell cycle. In summary, our results outline the backbone of a novel regulatory module that controls dormancy release and reactivation of growth in poplar.

PE0632: Forest Trees

A Poplar Receptor-like Kinase Mediates the Symbiotic Interaction between Plant and Mycorrhizal Fungi

Zhenzhen Qiao, Wellington Muchero and Jay Chen, Oak Ridge National Laboratory, Oak Ridge, TN

The soil-borne microbe/fungi can establish a mutualistic relationship with plant roots, to provide a large variety of nutrients to host plants in exchange of photosynthesized sugars. However, the molecular signal mediating the establishment of that relationship remains unclear. Our previous genetic mapping and resequencing data identified a receptor-like kinase coding gene *PtRLK1* deletion event in *Populus deltoides* associated with a lack of root colonization by ectomycorrhizal fungi *-Laccaria bicolor*. And further study showed the overexpression of *PtRLK1* could introduce the *L. bicolor* colonization in the *P.t.* x *P.d.* hybrids roots. Here, through introducing *PtRLK1* into *P. deltoides* and *P. tremula* x *alba* 717-1B4, we are able to show that the *PtRLK1* can help establishing the symbiotic relationship between non-host poplar plants and *L. bicolor*. In addition, our qPCR results showed the defense-related genes were oppositely regulated in the inoculated root sample of transgenic *P. deltoides* and 717-1B4 compared to the wild type, suggesting the over-expression of gene *PtRLK1*, to some extent, is able to switch gear for the defense mechanisms in poplar tree. Furthermore, we also over-expressed *PtRLK1* into perennial grass-switchgrass and annual crop grass-rice and observed that the *L. bicolor* is able to penetrate into the transgenic switchgrass root cortex and even to the vesicular tissues, but not into the wild type plant roots. We also observed a decline of biomass in the transgenic switchgrass. The metabolomic study on the *L. bicolor* inoculated transgenic switchgrass showed an increasing in N-containing metabolites and a decline in organic acids, sugars, and some phenolics like hydroxycinnamate conjugates, compared to the inoculated wild type switchgrass. To date, we have all RNA, metabolite, and phosphor-protein have been extracted from poplar, switchgrass and rice inoculated root samples and respect -omic analyses are undergoing.

PO0633: Forest Trees

Local Adaptation in *Populus trichocarpa* torr. & Gray

Hari Chhetri, West Virginia University, Morgantown, WV

Because of its rapid growth, hybrid vigor, broad geographic distribution, transformation potential, and the availability of tremendous genetic resources and wide phenotypic variation, *Populus* is a highly desirable species for biofuel production and other wood products. Understanding the genetic mechanisms underlying local adaptation is key for the sustainable management and domestication of forest trees like *Populus*. Here we report on the possible mechanisms underlying local adaption in *Populus trichocarpa* using whole genome re-sequencing, phenotypic and geo-climate data for 869 trees. First, we show that morphological and physiological traits are strongly correlated with the geo-climate variables of the source locations in *P. trichocarpa*. Second, using Genotype-Environment Association Analysis (GEA) and Redundancy Analysis (RDA) we identified several outlier loci that occur within and near genes related to important plant physiological functions and cuticular wax formation. A total of 32 genes were shared between RDA and GEA methods. Third, using RDA, we decomposed the among population variance of 869 trees into climate and geography. While climate and geography predictors together explained 7.6% of the total variation in the SNP matrix, climate alone explained 2.6% of the total variation. Partitioning the variance components in the response matrix of phenotypic traits into the explanatory matrices of SNPs, climate and geography, explanatory matrices altogether explained 22% of the total variation, whereas SNPs alone explained

3.9% of the total variation. These findings have important implications for developing management and conservation strategies and sustainability of forest resources in the face of climate change.

PE0634: Forest Trees

Genetic Diversity in *Populus trichocarpa* for Rare Variant Genetic Associations Querying 1,000+ Genomes

Anthony Piot, Université Laval, Québec, QC, Canada

Rare small genetic variants are found in very large number across genomes and can have important effects on coding sequences. As such, these rare variants hold great potential to explain a significant part of the missing heritability observed in most genetic association studies. Given their low frequency in populations ($MAF < 0.05$), however, rare variant identification requires dataset of large numbers of individuals. Moreover, rare variants need to be identified with confidence, as they can be confounded with sequencing errors. In this work, we use a filtered dataset of 1,014 pure *Populus trichocarpa* to identify rare and common small genetic variants across individual nuclear genomes. We compare variant calls between two software types and applied strict quality filters for improved genetic variant identification. Finally, we retain genetic variants that were identified by both variant callers, thus increasing calling confidence. We found a high genomic diversity in *P. trichocarpa*, with 7.4 million small genetic variants. Importantly, 358k non-synonymous and 25k nonsense variants were uncovered. GO enrichment analysis showed that genes with nonsense variant were enriched in functions related to wood formation. Using RNA-seq data, we further analyze the non-synonymous variants at the transcriptomic level in order to pinpoint genetic variants located in pseudogenes. We highlight the importance of genomic diversity and rare nonsense variants in explaining more of *P. trichocarpa*'s phenotypic variability in association genetics. The goal is to associate both rare and common alleles with poplar's wood quality traits to support selective breeding for an improved bioenergy feedstock.

PO0635: Forest Trees

Microevolutionary Signatures in the Sex Chromosomes and Organelles of *Populus trichocarpa*

David Macaya-Sanz¹, Chanaka Roshan Abeyratne¹, Ran Zhou¹, Jeremy Schmutz², Gerald Tuskan³ and Stephen DiFazio¹, (1)West Virginia University, Morgantown, WV, (2)HudsonAlpha Institute for Biotechnology, Huntsville, AL, (3)Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN

Sex chromosomes are cytological structures of evolutionary importance that have independently evolved and progressed in multiple branches across the tree of life. However, little is known about the forces and constraints that drive the population dynamics of sex chromosomes in plants. The genus *Populus* possesses sex chromosomes and is deemed as a model organism. In particular, *P. trichocarpa* (black cottonwood) has heterogametic males (XY system). We have used whole-genome sequencing (WGS) to identify and map the sex-determining region (Y-SDR) in the first assembled genome of a male cottonwood. We have called variants across the entire genome, including the X and Y chromosomes and the organelles, using WGS of 1200 wild accessions, comprising most of the natural range of this species. We have validated the calls and studied patterns of segregation and recombination using an additional WGS dataset of ca. 800 controlled cross progeny (derived from a full factorial cross between seven females and seven males). We have studied patterns of genetic diversity, population structure, genetic differentiation and evolutionary history for the autosomes, sex chromosomes, and organelles in this population. We have also identified signatures of haplotypic cooccurrence among these different elements to expose shared mechanisms of dispersal and imprints of coevolution and cytonuclear interactions. As expected, genotypic and coverage patterns in the Y-SDR are sufficient to predict sex. All of the genomic elements studied displayed considerable genetic diversity. As a matter of fact, both Y-SDR and organelles had sufficient resolution to accurately assign controlled cross progeny to the correct pedigree. They also exhibited stark but uncoupled geographic structure, and distinct long-term effective population sizes.

PE0636: Forest Trees

Expression Quantitative Trait Nucleotide Mapping in *Populus*

Jay Chen¹, Jin Zhang¹, Meng Xie¹, Yongil Yang¹, Kaijie Zheng¹, Lee Gunter¹, Sara S. Jawdy¹, Kerrie W. Barry², Jeremy Schmutz³, Gerald Tuskan⁴ and Wellington Muchero¹, (1)Oak Ridge National Laboratory, Oak Ridge, TN, (2)DOE Joint Genome Institute, Berkeley, CA, (3)HudsonAlpha Institute for Biotechnology, Huntsville, AL, (4)Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN

Populus is a model plant for studying woody perennial species and is being used for biomass production for a suite of industrial applications including biofuel and bioproduct production. To uncover the genetic regulatory landscape in *Populus*, we performed an expression quantitative trait nucleotide (eQTN) mapping enabled by the whole-genome resequencing and RNA-seq analysis of *P. trichocarpa* natural variants. A panel of >8.2 million single nucleotide polymorphisms (SNPs) and nucleotide insertions and deletions (InDels) were obtained from whole-genome resequencing of 917 unrelated individuals of *P. trichocarpa*. Transcriptome data from 390 leaf and 444 xylem samples were analyzed and revealed that 16,030 and 15,496 genes, respectively, exhibited significant expression variation across the population. Through genetic mapping, *cis*- and *trans*-eQTNs were identified and *trans*-eQTN analysis revealed multiple *trans*-eQTN hotspots that were significantly associated with the expression of more than 100 putative target genes. Furthermore, analysis of enriched transcription factor binding sites including *cis*-eQTNs revealed tissue-specific divergence. Combined with genome-wide association studies (GWAS) of trait phenotypes, the upstream regulators of these phenotype-associated genes and their regulatory network were identified. These analyses provide insight into the genetic regulatory mechanisms underlying complex traits.

The Center for Bioenergy Innovation is a U.S. Department of Energy Bioenergy Research Center supported by the Office of Biological and Environmental Research in the DOE Office of Science.

PO0637: Forest Trees

Growth-Defense Tradeoffs in North American Aspen (*Populus tremuloides*)

Jennifer F Riehl, University of Wisconsin-Madison, Madison, WI

Physiological tradeoffs in plant allocations to growth versus defense govern the ecology and evolution of tree-insect interactions and influence sustainable production of forest ecosystems. Despite the importance of plant defenses to forest health, we do not fully understand the genetic architecture of growth-defense trade-offs. *Populus* provides an ideal model system to study the genetic architecture of different resource allocation strategies, as phenolic defense compounds (e.g., phenolic glycosides) are negatively correlated with growth. Our study examined the genetic architecture underlying variation in key growth and chemical defense traits in a foundation forest tree species (*Populus tremuloides*) using genome-wide association analysis with complementary methods (multiple marker association and differential expression analyses). A large association mapping common garden of *P. tremuloides* was established in 2010 with four replicate blocks of genotypes (N = 515) collected from a north-south transect throughout Wisconsin, U.S.A. We evaluated a suite of relevant tree traits between 2014 and 2017. Sequence capture genotyping of the WisAsp population resulted in the discovery of ~115,000 SNPs. Budbreak and phenolic glycosides showed high broad-sense heritability, while growth traits showed low to moderate heritabilities. Single-marker GWA analyses have identified 13 candidate genes in both growth and defense traits, including a chorismate synthase gene associated with defense chemistry known to be involved in the biosynthesis of phenolic glycosides. Multiple-marker GWA analysis revealed that for all traits, no more than 35% of the variation explained by our SNP set was attributed to SNPs with larger effects, suggesting mostly polygenic control. Traditional single-marker GWA analyses are unlikely to detect influential genes for a given trait when the genetic architecture of the trait is polygenic with many genes contributing small to moderate effects. As result we are currently pursuing analyses that mediate this limitation. We are exploring the multi-marker GWA results further to identify candidate genes. Additionally, we are analyzing transcriptomic data from a subset of trees with extremely low and high levels of constituent phenolic glycosides to identify expressed genes involved in the strategy to invest more or less in defense and consequently more or less growth via differential expression analysis.

PE0638: Forest Trees

Elucidating Genetic Pathways of Sex Determination and Dimorphism in *Salix purpurea*

Brennan Hyden¹, Craig H. Carlson², Guanqiao Feng³, Matthew S. Olson³, Ran Zhou⁴, David Macaya-Sanz⁴, Stephen DiFazio⁴, Ken Keefover-Ring⁵ and Lawrence B. Smart², (1)Cornell University, Geneva,

NY, (2)Horticulture Section, School of Integrative Plant Science, Cornell University, Geneva, NY, (3)Texas Tech University, Lubbock, TX, (4)West Virginia University, Morgantown, WV, (5)University of Wisconsin-Madison, Madison, WI

Shrub willow (*Salix* section *Vetrix*), is a dioecious, short rotation coppice, bioenergy crop. While there are thousands of genes involved in sex dimorphism in shrub willow, there is an interest in mapping the gene networks associated with sex dimorphism in order to ultimately identify master regulator genes for sex. Here we report on expression QTL, differential expression analysis, and network analysis conducted in *Salix purpurea*, an economically important and model shrub willow species with a ZW system and a sex determining region (SDR) mapped to 6.73 Mb on Chr15W. RNA-Seq data were obtained from 90 males and 90 females and small RNA data from 22 males and 22 females in an F₂ population and mapped to the recent version 5.1 genome assembly. Mapped reads were subsequently utilized for eQTL, DESeq, and network analysis. These data were used to identify gene associations and networks strongly associated with sex, and to develop hypotheses regarding master regulator genes of sex. We present these sex-associated network modules and hypothesized pathways and mechanisms of sex determination and dimorphism via candidate master regulator genes. This study is the first transcriptome-wide network analysis in *Salix* floral tissue and represents a significant step towards understanding sex determination in the genus *Salix*. Results from this study can also provide valuable insight and knowledge to sex determination in the related genus *Populus*, as well as other dioecious plant species.

PO0639: Forest Trees

The Ancient Salicoid Genome Duplication Event: A Platform for Intra-Genome, Inter-Species and Inter-Genera Reconstruction of *de novo* Gene Evolution in *Populus trichocarpa*.

Timothy B. Yates¹, Kai Feng², Jin Zhang², Priya Ranjan², Paul E. Abraham², Lee Gunter², Sara S. Jawdy², Vasanth Singan³, Kerrie W. Barry⁴, Jeremy Schmutz⁵, Jay Chen², Gerald Tuskan⁶ and Wellington Muchero², (1)University of Tennessee, Knoxville, TN, (2)Oak Ridge National Laboratory, Oak Ridge, TN, (3)DOE Joint Genome Institute, Walnut Creek, CA, (4)DOE Joint Genome Institute, Berkeley, CA, (5)HudsonAlpha Institute for Biotechnology, Huntsville, AL, (6)Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN

Orphan genes lack sequence similarity to genes in other species and represent an important part of genome evolution as a novel source of genetic material. We identify 446 genes specific to *P. trichocarpa* with a small subset of the 446 orphan genes showing evidence of *de novo* gene evolution. *Populus* and its sister genera *Salix* are particularly well suited for the study of orphan gene evolution as a result of the Salicoid whole genome duplication (WGD) which resulted in highly syntenic sister chromosomal segments across the Salicaceae. We leverage this genomic feature to reconstruct *de novo* gene evolution from inter-genera, inter-species, and intra-genomic perspectives, by comparing the syntenic regions within *P. trichocarpa*, then *P. deltoides*, and finally *Salix purpurea*. Additionally, we also utilize the *Populus* Genome-wide association mapping panel (GWAS) population, a collection of 1,084 undomesticated genotypes to further understand the population genetics of orphan gene evolution. Furthermore, we use transcriptomics and proteomics to provide evidence of function for a large cohort of orphan genes. Overall, we provide new insights into the processes of *de novo* gene evolution in the context of a long-lived perennial tree.

PE0640: Forest Trees

Physiological and Transcriptomic Analysis of Sugar Maple under Drought Stress

Lungowe Mulozi, Kalpalatha Melmaiee, Krishnanand Kulkarni, Philip Bissiwu and Sathyanarayana Elavarthi, Delaware State University, Dover, DE

The sugar maple tree (*Acer saccharum*), native to Canada, is known for its hardwood and maple syrup production. Drought is one of the many environmental stresses that affect the growth and productivity of sugar maple trees; however, little about its response to the drought stress has been investigated so far. The objective of this study was to evaluate the response of sugar maple saplings to the drought stress through physiological analysis and identify differentially expressed genes (DEGs) through transcriptome sequencing. Three different groups of sugar maple saplings were subjected to drought stress over period of 21 days, and response was evaluated at three time points: 7, 14, and 21 days. The soil moisture content and physiological characteristics (simple ratio, NDVI, chlorophyll

content, and greenness index) of the drought-stressed plants as well as control plants were measured. For transcriptome analysis, leaf tissues were collected and RNA was extracted from the samples (in triplicates), and utilized for transcriptome sequencing using Illumina sequencer. Significant differences in the physiological parameters in drought-induced, and control plants were observed, thereby confirming the effect of drought on the sugar maple plants. Transcriptome sequencing results are expected to identify differentially expressed genes, which may further help understand the genes and pathways underlying the response of sugar maple trees to drought stress. The information generated in this study will form a great transcriptome resource and will be valuable for further study of molecular mechanisms of drought tolerance in sugar maple.

PO0641: Forest Trees

Sequencing and Phased Assembly of an *Eucalyptus urophylla* x *E. grandis* F₁ Hybrid and Parental Genomes

Anneri Lötter¹, Tuan A. Duong¹, Eshchar Mizrahi¹, Jill L. Wegrzyn² and Alexander A. Myburg¹,
(1)Department of Biochemistry, Genetics and Microbiology, Forestry and Agricultural Biotechnology Institute (FABI), University of Pretoria, Pretoria, South Africa, (2)Department of Ecology and Evolutionary Biology, University of Connecticut, Storrs, CT

Haplotype-based molecular breeding strategies may allow tree breeders to select desired allelic combinations in pure-species and hybrid genotypes. However, high-quality, phased genome assemblies are needed to allow identification of haplotype variation present within individuals. Long-read sequencing technologies enable new insights into haplotype variation in plant genomes, as longer read-lengths can span across homozygous regions and connect haplotypes for phased genome assembly. The aim of this study is to obtain high-quality, phased reference genomes for *Eucalyptus urophylla* and *E. grandis* by Nanopore sequencing and separate assembly of the two haplotypes in an *E. urophylla* x *E. grandis* F₁ hybrid. Towards this, Nanopore sequencing data obtained for two different high molecular weight DNA isolation methods were compared, resulting in 9.5 Gb and 2.3 Gb of MinION sequencing data passing QC. Using the combined data (17X genome coverage), 547 Mb of the genome was assembled into 6,124 contigs (contig N50 of 321 Kb). Next, we are expanding genome coverage to 100X coverage using PromethION long-reads. Illumina data will also be generated to allow phasing of the two haplotypes of the F₁ hybrid using a binning approach. Additionally, high-density SNP genetic linkage maps of both parents of the F₁ hybrid will be used to anchor scaffolds and improve chromosome-scale contiguity. These phased genome assemblies will be a first step towards generating a haplotype map for the parents of a large F₁ hybrid trial that will serve as reference for genome imputation and haplotype-based association mapping of growth and wood properties in the F₁ progeny.

PE0642: Forest Trees

Sweet Genomes: Sequencing, Assembling, and Annotating Three Maples

Susan McEvoy¹, Alexander Trouern-Trend¹, U. Uzay Sezen¹, Jill L. Wegrzyn¹ and Nate Swenson²,
(1)Department of Ecology and Evolutionary Biology, University of Connecticut, Storrs, CT,
(2)University of Maryland, College Park, MD

The reference genomes for *Acer negundo* (box elder), *Acer saccharum* (sugar maple), and *Acer griseum* (paperbark maple) represent the most comprehensive characterization of the 150 member *Acer* genus to date. Box elder is widely spread across North America and is commonly used as a fast growing ornamental in urban areas, while sugar maple, valuable for both its wood and sap, represents a narrower range in Northern portions of the Eastern and Central U.S. Sugar maple populations are declining and moving North in response to abiotic stressors, while the phylogenetically distant box elder has proved resilient to similar impacts. Paperbark is a panthocarpic species used as a landscaping tree in the U.S., though native to the Yunnan province in China where populations are endangered. These genomic resources will contribute to the relatively small collection of hardwood genomes sequenced to date, and provide a basis for investigations of their adaptive potential. The sequencing design consists of deep long read coverage (90x) from Nanopore and Pacific Biosciences SEQUEL, short reads from Illumina HiSeq (150bp PE), and Hi-C data (negundo, 100x; saccharum, 65x). These diploid, highly heterozygous trees have moderate genomes, estimated at 440Mbp 590Mbp, and 452Mbp respectively. Gene annotation combined existing and novel approaches to evaluate gene prediction methods, leveraging RNA-Seq data generated for all species. Genomic comparisons

among the four existing maples and other annotated land plants was used to identify putative expansions and contractions of gene families underlying the unique and shared biology of these species.

PO0643: Forest Trees

Development of 20K Genotyping Array and Construction of High-Density Genetic Map of Hybrid Larch (*Larix gmelinii* var. *japonica* × *Larix kaempferi*)

Kentaro Mishima¹, Hideki Hirakawa², Tomonori Hirao³, Taiichi Iki⁴, Yoko Fukuda⁵, Akira Tamura⁶, Manabu Kurita⁷ and Makoto Takahashi¹, (1)Forest Tree Breeding Center, Forestry and Forest Products Research Institute, Hitachi, Ibaraki, Japan, (2)Kazusa DNA Research Institute, Kisarazu, Chiba, Japan, (3)Forest Bio-Research Center, Forestry and Forest Products Research Institute, Hitachi, Ibaraki, Japan, (4)Tohoku Regional Breeding Office, Forest Tree Breeding Center, Forestry and Forest Products Research Institute, Takizawa, Iwate, Japan, (5)Hokkaido Breeding Office for FTBC, Forestry and Forest Products Research Institute, Ebetsu, Hokkaido, Japan, (6)Forest Breeding Center, Forestry and Forest Products Institute, Hitachi, Japan, (7)Kyushu Regional Breeding Office, Forest Tree Breeding Center, FFPRI, Koshi, Kumamoto, Japan

Japanese larch (*Larix kaempferi*) is a deciduous, coniferous species that naturally grows cool-temperature forest from central to northern Honshu, and it is endemic in Japan. The species is one of the most important forestry tree species in Japan. Kurile larch (*Larix gmelinii* var. *japonica*) is a variety of Dahurian larch (*Larix gmelinii*), distributed in Chishima islands (Kuril islands) and Sakhalin. Because the hybrid larch (*Larix gmelinii* var. *japonica* × *Larix kaempferi*) shows heterosis (rapid juvenile growth, high resistance to various pests and diseases), eleven million seedlings/year are supplied as planting material for afforestation in Hokkaido region.

We identified 79,832 isoforms as reference sequence by PacBio RSII. For SNP discovery, resequencing to reference sequences was performed for several tissue from five plus-trees, and 10Gb short read sequences was obtained from each library using the illumina HiSeq 2500. Approximately 20k SNPs were used for development of Axiom genotyping system. We constructed a high-density linkage map using the hybrid larch F₁ population (*Larix gmelinii* var. *japonica* × *Larix kaempferi*). About 4k SNPs was polymorphic between the parents. The linkage map consisted of 12 linkage groups of 1,727 loci (average marker distance: 1.05 loci/cM) spanning 1809.9cM.

PE0644: Forest Trees

Extraction of Differential Expressed Genes in High Temperature Environmental Response of Japanese Cedar, *Cryptomeria japonica*

Soichiro Nagano, Mine Nose, Yuya Takashima, Michinari Matsushita, Kentaro Mishima and Tomonori Hirao, Forest Tree Breeding Center, Forestry and Forest Products Research Institute, Hitachi, Ibaraki, Japan

With global warming, forest trees are exposing to high-temperature environments. We have conducted comprehensive gene expression analysis in order to clarify the response ability of *Cryptomeria japonica* to high-temperature environments and to verify the differences between strains. Cuttage seedling of four strains were grown under three conditions: control, daytime high-temperature, and nighttime high-temperature, and the current needles were harvested at 0 d, 7 d, and 14 d of the high-temperature treatment. Total RNA was extracted from the current needles, an RNA-Seq library was prepared, and 101 bp PE reads were obtained with Illumina HiSeq4000. The sequenced reads were mapped to the reference gene sequence of *C. japonica*, and the expression level for each gene was calculated. Although, the gene expression characterizing the strain was more remarkable than the high-temperature response of each strain, the expression of genes involved in membrane lipid and HSP tended to increase and the gene involved in cell division tended to decrease in the high-temperature environment. This study suggests that the high temperature environmental response of *C. japonica* can be evaluated with gene expression analysis.

PO0645: Forest Trees

Mānuka and Its Microbiome: A Multi ‘-Omics Approach to Look into an Unique Plant

Cecilia H. Deng¹, Ella Grierson², Elena Hilario¹, Lindy Guo¹, Kathy Schwinn², Di Brewster¹, Janine Cooney¹, Dan Jones³, Hayley Ridgway¹, David Chagné² and Amali H. Thrimawithana¹, (1)The New Zealand Institute for Plant & Food Research Limited, Auckland, New Zealand, (2)The New Zealand Institute for Plant & Food Research Limited, Palmerston North, New Zealand, (3)Plant & Food Research, Auckland, New Zealand

Mānuka, or *Leptospermum scoparium*, is a tree native to New Zealand and Australia that is renowned not only for its essential oils but also for the high value honey derived from its floral nectar. The study of microbiome impact on mānuka is still in its infancy, with only a few studies looking at bacterial signatures or focusing on particular species. Microbiomes play critical roles in plant growth and health. Their composition and abundance may influence a plant's nutrient intake, provide protection against biotic and abiotic stress, and may assist the host to adapt to climate changes. To gain insights to the mānuka microbiome and to investigate the coevolution of host-microbe interactions in mānuka, we are utilizing a multi '-omics approach which integrates data from the mānuka genome, tissue-specific transcriptome, proteome and cultureome. Diversity in bacterial and fungal composition has been observed across assorted types of tissues in a mānuka tree, with clear differences seen between mature and young tissues. This study initiates our understanding of the plant-microbiome interaction, and provides resources beneficial to not only the research community but also the mānuka industries. Mānuka has special cultural significance to the indigenous people of New Zealand. The data developed from this work will be shared with the scientific community after due consideration that Māori core cultural values are respected, and that intended use of the knowledge will have positive impacts on Māori social, cultural, environmental and economic well-being.

PE0646: Forest Trees

Exploring the Role of Differential Gene Regulation on *Ginkgo biloba* Morphology

Austin Nicole Cauwels, Cal State East Bay, Hayward, CA and Ana Almeida, California State University East Bay, Hayward, CA

Ginkgo biloba is a species of gymnosperm that has existed for 200 million years, earning it the title 'living fossil'. Compared to other gymnosperms *Ginkgo biloba* exhibits an interesting leaf morphology. In this project, RNA was extracted from the flat laminar blade and the radially symmetric petiole in order to process the RNA transcriptome with computational models. Through this Abaxial-Adaxial Gene Regulatory Network (Polarity GRN) properties were studied to observe plant development. Genomic sequencing data was used to perform de novo RNA sequence assembly. Differential expression genes were visualized with a heatmap and 147 differentially expressed genes were found to be differentially expressed. Of these genes, 62.6% of genes were upregulated in the petiole and 37.6% were upregulated in the blade. qPCR data of RNA extracted from primordial blades and petioles will be presented alongside the bioinformatics generated from *Ginkgo biloba* transcriptome.

PO0647: Forest Trees

Assembly and Annotation of the American Beech (*Fagus grandifolia*) Genome

Akriti Bhattarai¹, Andrew Starovoitov¹, Sunny Sarker¹, Nate Swenson² and Jill L. Wegrzyn¹, (1)Department of Ecology and Evolutionary Biology, University of Connecticut, Storrs, CT, (2)University of Maryland, College Park, MD

The native range of the North American hardwood, American beech (*Fagus grandifolia*), stretches from Nova Scotia to northern Florida and as far west as eastern Texas. This highly shade-tolerant and slow growing species is found in many northern hardwood forests. In the northeastern portion of its range, it is increasingly susceptible to beech bark disease instigated by the introduced beech scale insect. The insect damage leads to subsequent fungal attack, stem cankers, and irreversible stem damage, often leading to mortality via *beech snap*. To assist in population level studies aimed at identifying natural resistance, the American beech genome, estimated at 420 Mb, was sequenced using a combination of Illumina HiSeq (250bp PE) short reads (89X coverage) and Oxford Nanopore PromethION long reads (165X). The genome was assembled with Shasta (v.0.2.0) to obtain a total of 3,085 scaffolds representing 456 Mb with an N50 of 599 Kb. BUSCO assessment via the viridiplantae database reflected 75% of the complete and single-copy proteins prior to polishing. The genome was subsequently filtered for contaminant reads, scaffolds < 3Kb, and polished to produce a final assembly. The polished BUSCO assessment

identifies the majority of the putative single-copy proteins. Phase genomics Hi-C libraries were prepared to further scaffold the assembly. RNA-Seq reads derived from leaf tissue were aligned to the reference genome to provide evidence for the Braker (v2.1.2) genome annotation package. The final annotation was functionally annotated with EnTAP and filtered for high quality gene models via gFACs.

PE0648: Forest Trees

Computational Identification of Conserved miRNAs and Their Putative Target Genes in *Juglans regia* and *J. microcarpa*

Le Wang, Tingting Zhu, Karin R. Deal, Jan Dvorak and Ming-Cheng Luo, Department of Plant Sciences, University of California, Davis, Davis, CA

MicroRNAs (miRNAs), small non-coding RNAs with 20 to 24 nucleotides, play important roles in the post-transcriptional regulation of protein-coding genes in plants. High quality reference genome assemblies for *Juglans regia* and *J. microcarpa* make it possible for us to identify miRNAs using a homology-based approach and characterize their distribution on the chromosomes of both species. We identified 39 and 40 conserved miRNAs belonging to 27 and 26 miRNA families in *J. regia* and *J. microcarpa*, respectively. There is no significant difference in either the quantity or category of identified miRNAs between the two species. An asymmetric distribution of miRNA precursors between the dominant and subdominant chromosomes in both species was detected. We also predicted 325 and 316 potential target genes for these miRNAs in *J. regia* and *J. microcarpa*, respectively. Of these target genes, 12 and 18 resistance gene analogues and 56 and 54 transcript factors were found in *J. regia* and *J. microcarpa*, respectively. Functional annotation demonstrated that miRNAs regulate carbohydrate metabolism, environmental information processing, and signaling and cellular processing in *J. regia* and *J. microcarpa*.

PO0649: Fruit Species

BIMS (Breeding Information Management System) for Efficient Management and Analysis of Breeding Data

Sook Jung¹, Taein Lee², Chun-Huai Cheng², Ksenija Gasic³, B. Todd Campbell⁴, Jodi L. Humann¹, Jing Yu¹, Heidi Hough¹ and Dorrie Main⁵, (1)Washington State University, Pullman, WA, (2)Washington State University, Pullman, Pullman, WA, (3)Clemson University, Clemson, SC, (4)USDA-ARS, FLORENCE, SC, (5)Department of Horticulture, Washington State University, Pullman, WA

Breeding programs produce a large amount of data that require efficient management systems to keep track of performance, pedigree, geographical and image-based data as well as genotyping data. The integration of breeding data with publicly available genomic and genetic data, as well as the integration of each breeder's own genotypic and phenotypic data in a database enhances genetic understanding of important traits and maximizes the marker-assisted breeding utility by breeders and allied scientists. We report the progress on BIMS which we have implemented in the Genome Database for Rosaceae, CottonGEN Citrus Genome Database, Pulse Crop Database and the Genome Database for Vaccinium. BIMS allows individual breeders to integrate their phenotypic and genotypic data with public genomic and genetic data and at the same time have complete control of their own breeding data and access to tools such as data import/export, data analysis and data archive. BIMS incorporates the use of the Android App Field Book, an open-source software for phones and tablets which allows breeders to replace hard-copy field books, thus alleviating the possibility of transcription errors while providing faster access to the collected data. The use of Field Book and BIMS promotes the use and development of standard trait descriptors and metadata as well. New functionality includes searching/loading SNP genotype data and haplotype data, cross search and bulk data editing.

PE0650: Fruit Species

De Novo Phased Assembly of the *Vitis riparia* Grape Genome

Pierre-François Bert, ISVV - UMR 1287 EGFV - INRA - University of Bordeaux, Villenave d'Ornon, France and **Pierre-François BERT**, INRA EGFV UMR 1287, VILLENAVE D'ORNON, France

Grapevine is one of the most important fruit species in the world. In order to better understand genetic basis of traits variation and facilitate the breeding of new genotypes, we sequenced, assembled, and annotated the genome of the American native *Vitis riparia*, one of the main species used worldwide for rootstock and scion breeding. A total of 164 Gb raw DNA reads were obtained from *Vitis riparia* resulting in a 225X depth of coverage. We generated a genome assembly of the *V. riparia* grape *de novo* using the PacBio long-reads that was phased with the 10x Genomics Chromium linked-reads. At the chromosome level, a 500 Mb genome was generated with a scaffold N50 size of 1 Mb. More than 34% of the whole genome were identified as repeat sequences, and 37,207 protein-coding genes were predicted. This genome assembly sets the stage for comparative genomic analysis of the diversification and adaptation of grapevine and will provide a solid resource for further genetic analysis and breeding of this economically important species.

PO0651: Fruit Species

Improving Performance of Variant Calling from Genome Resequencing and Insights into Genes Controlling Seedlessness Trait in the Genus *Vitis*

Myung-Shin Kim¹, Ji Hong Kim², Youn Young Hur³ and **Soon-Chun Jeong²**, (1)Department of Plant Science, Plant Genomics and Breeding Institute and Interdisciplinary Program in Agricultural Genomics, Seoul National University, Seoul, Korea, Republic of (South), (2)Korea Research Institute of Bioscience and Biotechnology, Chungbuk, South Korea, (3)National Institute of Horticultural and Herbal Science, Rural Development Administration, Wanju, Korea, Republic of (South)

The seedlessness of grape derived from stenospermocarpy is one of the most prized traits of table or raisin grapes. It is controlled by a complex genetic system containing one major dominant gene and multiple minor recessive genes. Here, we collected dense variation data from high-depth resequencing data of seeded, seedless, and wild relative grape genomes sequenced to > 37x mean depth. Variant calls were made using a modified variant calling pipeline that was sufficient for highly diverse interspecific grape accessions. The modified pipeline enabled us to call several million more variants than the commonly recommended pipeline. The quality was validated by Sanger sequencing data and subsequently supported by the genetic population structure and the phylogenetic tree constructed using the obtained variation data, results of which were generally consistent with known pedigree and taxonomic classifications. Variation data enabled us to confirm a major dominant QTL (quantitative trait loci) and identify minor recessive QTL for seedlessness. Incidentally, we found that grape cultivar Rizamat contains an ancestral chromosomal region of the major QTL in Sultanina, a predominant seedlessness donor cultivar. Furthermore, we predicted new candidate causal genes including *Vitvi01g00455*, *Vitvi08g01528*, and *Vitvi18g01237* associated with the minor seedless-regulating QTL, which showed high homology with genes that regulate seed development in *Arabidopsis*. This study provides fundamental insights relevant to variant calling from genome resequencing data of diverse interspecific hybrid germplasms such as grape and will accelerate future efforts aimed at crop improvement

PE0652: Fruit Species

Heavy Metals Content in Contrasting Local European Grapevine Genotypes

Georgiana D. Gabur¹, Carmen Teodosiu¹ and Valeriu V. Cotea², (1)"Gheorghe Asachi" Technical University, Iasi, Romania, (2)"Ion Ionescu de la Brad" University of Agricultural Sciences and Veterinary Medicine, Iasi, Romania

Heavy metals content needs to be controlled regularly and represents a prerequisite for safe grape consumption and human health. Genotypic specific Fe ions accumulation in grapes may be influenced by a wide range of factors, including the variety, nature of soils or the chemical treatments applied during vine growing management. Regarding the contamination sources, in the winemaking practice, Fe is also called „technological iron” and is distinguishing itself from the „physiological iron” originated in wine by natural physical-chemical reactions at biological stages of grape genotypes. The current study aim is to determine the concentration of iron (Fe) ions contaminants in wines processed from contrasting vine cultivars. Analysis of Fe concentration was done using a ContrAA 800 G high-resolution continuum source atomic absorption spectrometer (HR-CS AAS) equipped with graphite furnace. Commercial white and red wines varieties (Grasa de Cotnari and Feteasca neagra) from Cotnari area (Romania) were analyzed. Results indicated that Fe content was higher in red wines samples than in the white wines, due to the prolonged contact of the must with pomace. Moreover, derived differences can be detected and

interactions with environmental factors increase these differences. As an active compound in the processes of oxidation reduction, Fe plays an important part in wine evolution, especially during the aging phase. Genotypes screened in this study can contribute to future grapevine breeding purposes and at the pursuit to make safe and high-quality wines.

PO0653: Fruit Species

Molecular Evaluation of Vitality and Survival Rate of Grape Seedlings at Dormant Stage: A Step Toward Molecular Farming

Maazullah Nasim¹, Ting Zheng¹, Xiangpeng Leng², Haifeng Jia¹ and **Jingguo Fang¹**, (1)Nanjing Agricultural University, Nanjing, China, (2)Qingdao Agricultural University, Qingdao, China

Vitality and survival rate of grape seedlings are crucial factors affecting quality of vineyards. However, there is no comprehensive study describing some powerful evaluation of grapevine seedlings' vitality and survival rate at dormant stage. House Keeping Genes (HKGs) are involved in basic cellular functions and its expression level indicates vitality and survival of the plant. Therefore, the aim of this study was to investigate the application possibility of molecular information to evaluate dormant seedling vitality and survival rate prior planting. The expression level of six HKGs in buds of four parts of tetraploid Kyoho grape (*Vitis labruscana*: *V. labrusca* x *V. vinifera*) were detected by (Sq.) RT-PCR and qRT-PCR. The results revealed that higher expression of HKGs is strongly linked to high vitality and survival rate, low expression was associated to low vitality, Lower expression was significantly associated with low vitality and low survival rate, no survival of seedlings were seen with no expression of HKGs. DNA and RNA quality can roughly determine seedlings quality. Survival rate of the seedlings produced in Juxian-Shandong, Laixi-Shandong, Huailai-Hebei, Suizhong-Liaoning, Changli-Hebei, Guanxian-Shandong, and Zhangjiagang-Jiangsu province companies' were 100%, 100%, 100%, 100%, 100%, 87.77% and 93.33% respectively. In conclusion, the results suggest that, utilization of molecular techniques will be unique, rapid and potential approach for promoting gene information in seedling vitality measurement and would be a great method for evaluating the dormant grape seedling survival rate.

PE0654: Fruit Species

Selection across NBS-Encoding Genes during Apple Domestication

Jugpreet Singh¹, Manyi Sun², Weiha Zhang³, Liao Liao⁴, Jun Wu², Yuepeng Han⁴ and Awais Khan¹, (1)Plant Pathology and Plant-Microbe Biology, Cornell University, Geneva, NY, (2)Nanjing Agricultural University, Nanjing, China, (3)Huazhong Agriculture University, Wuhan, China, (4)Wuhan Botanic Garden of Chinese Academy of Sciences, Wuhan, China

Domestication has altered several phenotypic and molecular responses in plants. Previous studies have shown that domesticated plants exhibit more sensitivity to various pathogens, and pathogens may acquire high virulence on cultivated plants than their wild progenitors. These observations suggest that domestication has altered the host response through selection of disease-related genes and pathways. We used NBS-encoding gene family as a model to study the disease-related selection effects during the apple domestication. Genetic diversity and selection across NBS-encoding genes were analyzed using whole-genome resequencing data from a large set of domesticated (*Malus x domestica*; MD) and *Malus sieversii* (MS) accessions. Total 501 NBS-encoding genes were detected in the Golden Delicious double haploid (GDDH) apple genome. Proximal (43.3%) duplication events were mostly associated with the expansion of NBS-encoding genes in apple, which also led to their clustering and non-random distribution across the genome. The average nucleotide diversity across NBS-encoding genes was slightly higher in the domesticated apples (4.98×10^{-3}) than the wild accessions (4.71×10^{-3}). Also, genome-wide statistics of TajimaD (MD – 1.16; MS – 0.62) and fixation index (Fst; MD/MS – 0.16) suggest moderate differentiation between NBS-encoding genes from *Malus x domestica* and *Malus sieversii*. However, 1 kilobase genomic-window analysis had identified many NBS-encoding genes with strong selection signatures during apple domestication. Studying these genes can provide opportunities to understand disease susceptibility and host-pathosystem in *Malus x domestica* and *Malus sieversii*.

PO0655: Fruit Species

Pooled Genome Sequencing Analysis of Genetic Dwarf in Apple

Laura E. Dougherty, Susan K. Brown and Kenong Xu, Horticulture Section, School of Integrative Plant Science, Cornell AgriTech, Cornell University, Geneva, NY

In commercial apple production, trees are usually grafted on dwarfing rootstocks to control scion height and vigor. Genetic studies on apple rootstock dwarfing effect have identified important quantitative trait loci (QTLs) *Dw1*, *Dw2*, and *Dw3* (on chromosomes 5, 11 and 13, respectively) and their promising candidate genes. To understand the genetic control of dwarf in apple scion, we investigated dwarf in an own rooted F₁ population of 365 seedling trees, among which dwarf appeared to segregate as a recessive trait. Using a pooled genome sequencing approach, 'Tall' and 'Dwarf' genomic DNA pools were created based on tree height measurements in multiple years. Deep sequencing of the pooled genomes and subsequent detection and segregation analysis of single nucleotide variants (SNVs) that are present in both pools led to the discovery and confirmation of two QTLs linked to genetic dwarf on chromosomes 1 and 8, respectively. These findings indicate that the genetic mechanisms responsible for dwarf in this apple scion population are different from those involved in rootstock conferred dwarfing.

PE0656: Fruit Species

A Dense SNP-Based Genetic Linkage Map of Pear (*Pyrus* spp.) Anchoring SSR and Indel Markers

Yumi Kim¹, Hyeondae Han¹, Keumsun Kim¹, Sewon Oh¹, Jin-Ho Eom¹, Dongyun Lee¹, Yoon-Kyeong Kim², Do-Yeon Cho¹ and Daeil Kim¹, (1)Chungbuk National University, Cheongju, South Korea, (2)National Institute of Horticultural & Herbal Science, Naju, South Korea

Microarray is an effective technology for genotyping of single nucleotide polymorphisms (SNPs) and Axiom 70K Pear SNP array has been developed in pear (*Pyrus* spp.). In this study, a dense genetic linkage map was constructed by additionally anchoring the SNPs derived from the 70K Pear array into previously constructed genetic linkage map of 'Whangkeumbae' (*P. pyrifolia*) × 'Minibae' (*P.* hybrid). The genetic linkage map was constructed using 8,662 array-SNPs, the previously developed GBS-SNP, insertion/deletion (InDel), and simple sequence repeat (SSR) markers. Linkage analysis and visualization were carried out using JoinMap 5.0. and MapChart, respectively. The integrated genetic linkage map consisted of 1,197 loci including 454 array-SNP, 710 GBS-SNP, 17 InDel, and 16 SSR markers in 17 linkage groups (LGs) with a total length of 2,298 centi-Morgan (cM) and with an average marker interval of 1.9 cM. The number of markers on each LG ranged from 50 (LG13) to 106 (LG15) and the length of each LG ranged from 97.8 (LG4) to 190.9 cM (LG11). Compare to the previously constructed genetic linkage map, a total length of the integrated genetic linkage map was increased and its average marker interval was more denser than GBS-SNP based genetic linkage map. Because this integrated map could represent the physical location of mapped markers, it would be a useful tool for identification of quantitative trait loci, furthermore, it could be used for fine mapping of trait of interest.

PO0657: Fruit Species

Identification and Investigation of Ancient Subgenome Evolutionary Fate in Pear (*Pyrus bretschneideri* Rehd.)

Qionghou Li, Xin Qiao and Shaoling Zhang, Centre of Pear Engineering Technology Research, State Key Laboratory of Crop Genetics and Germplasm Enhancement, Nanjing Agricultural University, Nanjing 210095, China, Nanjing, China

Many species have undergone one or more whole-genome duplication events (WGD; as known as paleopolyploidization), resulting genome size expansion. Genome fractionation (also known as diploidization) often occurs after WGD events. Many paleo-allopolyploids has been found have biased fractionation, but biased fractionation cannot be found in paleo-autopolyploids even in some paleo-allopolyploids. However, pear two ancient subgenome evolutionary fate is still confused, since pear also undergone a recent WGD event (~ 30Mya). Here, we not only identified the two paleo-subgenomes in pear using peach (*Prunus persica*) as outgroup, but calculated gene loss rate, synonymous and nonsynonymous substitution rates, expression levels, and DNA methylation level between these two subgenomes. The phenomenon of fractionation bias also absents in pear genome and the two subgenomes have similar Ka, Ks and Ka/Ks ratio, that is, the two subgenomes evolved at similar evolutionary rates. For expression levels and DNA methylation levels, we also found no bias between two subgenomes. However, we found that the singleton genes and homeologous genes within each subgenome showed divergent evolutionary pattern in

selective constraints, expression and epigenetic modification. The results of this study provide insights into the subgenome evolution following paleopolyploidization in pear and other plants.

PE0658: Fruit Species

Identification of Unique Mitochondrial Genome Organization in the Descendants of 'Niitaka' (*Pyrus pyrifolia*)

Daeil Kim¹, Hyeondae Han¹, Keumsun Kim¹, Sewon Oh¹, Yumi Kim¹, Jihyun Oh¹, Jin-Ho Eom¹, Do-Yeon Cho¹ and Hae Won Jung², (1)Chungbuk National University, Cheongju, South Korea, (2)National Institute of Horticultural & Herbal Science, Naju, South Korea

Although cytoplasmic male sterility (CMS) of pear was reported in 1976, its molecular evidence has not been provided. We evaluated male sterility in the descendants of 'Niitaka' (*Pyrus pyrifolia*) and observed maternal inheritance of male sterility. We compared mitochondrial genome structures of four pear accessions; two were descendants of 'Niitaka' and were male sterile pears, others were male fertile that were not derived from 'Niitaka'. We identified that a unique mitochondrial genome organization of 'Niitaka' is located at downstream of *cox3* and upstream of *atp8*, which have been frequently associated with the formation of CMS-causing genes in various crops. Maternal inheritance of male sterility in the descendants of 'Niitaka' as well as identification of sequences of origins unique to 'Niitaka' strongly indicate the CMS inheritance of male sterility in *Pyrus* spp. The molecular markers (CBpMtd03 and CBpMtd07) developed will be used to precisely identify the male-sterile cytoplasm.

PO0659: Fruit Species

Comparative Analysis of the Volatile Organic Compounds in Mature Fruits of 12 Occidental Pear (*Pyrus communis* L.) Cultivars

Yangyang Chen, Nanjing Agricultural University, Nanjing 210095, China, Nanjing, China and Shaoling Zhang, Centre of Pear Engineering Technology Research, State Key Laboratory of Crop Genetics and Germplasm Enhancement, Nanjing Agricultural University, Nanjing 210095, China, Nanjing, China

Fruit aroma is one of the most important attributes of fruit quality. In this study, HS-SPME and GC-MS were used to detect aromatic components of matured fruit of 12 Occidental pear cultivars. A total of 335 volatile organic compounds were identified, including esters, alcohols, alkanes, acids, ketones, terpenes and aldehydes. The highest concentration of total aroma was found in 'Alexandrine Douillard' (18730 µg/kg), whereas the lowest total concentration was in 'Bartlett-Max Red' (330 µg/kg). Based on the ABC's of Perfumery System, the 12 pear cultivars were divided into two groups, in which 'La France', 'Abate Fetel', 'Bartlett', 'Beurre Bosc', 'Alexandrine Douillard', 'Doctor Jules Guyot' and 'Yubileen Dar' were classified into the fruit scent type (Group one); and 'Butirra Rosata Morettini', 'beurré Hardy', 'Bartlett-Max Red', 'Clapp Favorite' and 'Red Clapp Favorite' were classified into the aliphatic scent type (Group two).

PE0660: Fruit Species

Comparison of Transcriptome Profiles Between Resistant and Susceptible Pears in Response to Infection of *Venturia nashicola*

Sewon Oh¹, Keumsun Kim¹, Hyeondae Han¹, Yumi Kim¹, Dongyun Lee¹, Jihyun Oh¹, Kyungho Won² and Daeil Kim¹, (1)Chungbuk National University, Cheongju, South Korea, (2)National Institute of Horticultural Herbal Science, Naju-si, South Korea

For the identification of resistance mechanisms against pear scab (*Venturia nashicola*), transcriptome sequencing was performed using scab resistant and susceptible cultivars. According to the infection behavior of *V. nashicola*, leaves of the two cultivars were collected at 0, 6, 12, 48, and 72 hours after inoculation. A total of 4,744 and 6,506 differentially expressed genes (DEGs) were detected in scab resistant and susceptible cultivars, respectively. The correlation analysis and principal component analysis indicated that the expression time of defense-related genes could affect resistance to scab rather than genotype. Functional annotation of DEGs into gene ontology (GO) database showed that the top-3 GO terms in three main GO categories (biological process, cellular component, molecular function) contained defense-related genes encoding pathogenesis-related (PR) proteins, protein kinases,

and transcription factors. According to the clustering analysis, we could confirm that those defense-related DEGs were rapidly up-regulated in scab resistant cultivar. Particularly, two transcription factors (*ethylene-responsive transcription factor* and *WRKY*), *PR-10* (*major allergen Pru av 1*), and *receptor-like protein* were rapidly up-regulated and displayed higher gene expression levels in scab resistant cultivar. These results suggest that the fast recognition of *V. nashicola* and rapid up-regulation of defense-related genes confer resistance to scab.

PO0661: Fruit Species

Comparative Transcriptome Analysis in ‘Passe Crassane’ Pear Fruit Exposed to Low Temperature or Propylene

Takumi Otsuki, Oscar Witere Mitalo, Yasuaki Tosa, Yuki Kondo, Azimullah Azimi, Abdul Hakim Kazimi, Takashi Akagi, Koichiro Ushijima and Yasutaka Kubo, Okayama University, Okayama, Japan

European pear responds to low temperature (LT) treatments by inducing ethylene, leading to ripening. However, it is unclear to what extent this is the result of LT alone or LT-induced ethylene. In this experiment, we followed the physiological and molecular responses of ‘Passe Crassane’ pears to LT and the ethylene analogue, propylene. Fruit at 20 °C treated with propylene softened within 9–10 d, with little changes in endogenous ethylene. By contrast, LT-treated fruit (0 °C and 5 °C for 42 d) produced large amounts of ethylene, and rapidly softened after being transferred to 20 °C. From transcriptome analysis, we identified 437 differentially expressed genes (DEGs) between propylene-treated and control fruit, which were further augmented by LT treatment. While, the expression of 763 DEGs between 5 °C vs. 20 °C was not significantly affected by propylene treatment in non-LT fruit. To examine LT-induced and ethylene induced pathways separately during chilling, the responses of LT-induced DEGs to 1-methylcyclopropene (1-MCP), an ethylene action inhibitor, were assessed. Among the 763 LT-induced DEGs, 1-MCP treatment disrupted the expression of 390 DEGs, indicating that they were regulated by 1 LT-induced ethylene. Intriguingly, 373 DEGs including transcription factor-related genes were unaffected by 1-MCP treatment, and thus, likely to be influenced by LT alone. Based on these results, the potential role of these LT-specific genes and pathways as a key factor modulating changes in ethylene and responsiveness leading to ripening in European pears is discussed.

PE0662: Fruit Species

Functional Characterization and Mechanism Analysis of PbrMYB5 Gene from *Pyrus betulaefolia*

Caihua Xing¹, Shaoling Zhang² and Xiaosan Huang², (1)Nanjing Agricultural university, Nanjing, China, (2)Nanjing Agricultural University, Nanjing, China

Dehydroascorbate reductase (DHAR) plays an important role in plant stress response, but DHAR currently has little understanding of the transcriptional regulation of abiotic stress. In this study, a novel transcription factor of the R2R3 type MYB was isolated from *Pyrus betulaefolia* by yeast single hybrid technique and named PbrMYB5. PbrMYB5 was localized in the nucleus and could specifically bind to the PbrDHAR2 promoter. PbrMYB5 increased its expression under low temperature and salt stress, but its response to drought was not strong. Overexpression of PbrMYB5 in tobacco increased its cold tolerance, while silencing the gene by VIGS technology in *Pyrus betulaefolia* resulted in its sensitivity to low temperatures. Compared with the wild type, the expression level of NtDHAR2 in transgenic tobacco was higher, and the accumulation of ascorbic acid was larger. In the *Pyrus betulaefolia* strain in which the Virus-induced gene silencing of PbrMYB5, down-regulated PbrDHAR2 ascorbic acid content and increased sensitivity to low temperature. Taken together, physiological and molecular identification showed that the gene has cold-resistant function, and its cold resistant mechanism is that PbrMYB5 can directly regulate the synthesis of ascorbic acid by specifically binding to the PbrDHAR2 gene promoter.

PO0663: Fruit Species

Analysis of Genetic Variability of MYB10 Genes in Japanese Plum and Its Association with Fruit Skin Colour

Maria Jose Aranzana, IRTA-CRAG, Barcelona, Spain and Arnau Fiol, CRAG, Cerdanyola del Vallès (Barcelona), Spain

Red coloration in plant organs and fruits is caused by the accumulation of anthocyanins, which are water-soluble pigments generally associated with healthy aspects. In the species of the *Rosaceae* family (including pip and stone fruits), such accumulation is led by the expression of R2R3-MYB10 transcription factor genes. In this work we aim at studying the role of these genes in determining fruit colour variability in Japanese plum, a stone fruit member of the *Prunus* genus (*P. salicina*) within the *Rosaceae* family, which varieties show a wide spectrum of fruit skin colours, ranging from pale green to dark purple, including yellow and red hues. Genomic tools in Japanese plum are still limited, being peach the closest model species most extensively studied. In peach, MYB10.1 gene is candidate for skin coloration; this gene clusters in LG3 with other MYB10 homologous genes (MYB10.2 and MYB10.3). Here we explore the genetic variability in the LG3 MYB10 gene cluster in a panel of commercial varieties, advanced breeding lines and progenies of Japanese plum and its association with fruit skin colour. Whole-genome Illumina reads (27x) from a red and a yellow variety mapped against peach and other *Prunus* genomes available in databases, allele cloning and phylogenetic analysis provide some light towards the understanding fruit skin colour variability in Japanese Plum.

PE0664: Fruit Species

Combining Genome-Wide Association Analyses and QTL Detection to Identify Loci Controlling Phenology-Related and Fruit Quality Traits in Sweet Cherry

Camille Branchereau¹, Armel Donkpegnan¹, Teresa Barreneche¹, Loïck Le Dantec², Anton Ivancic³, Daniela Giovannini⁴, Gregorio López-Ortega⁵, Federico García-Montiel⁵, José Quero-García¹ and **Elisabeth Dirlewanger¹**, (1)UMR 1332 BFP, INRA, Université de Bordeaux, Villenave d'Ornon, France, (2)Univ. Bordeaux, Villenave d'Ornon, France, (3)Faculty of Agriculture and Life Sciences, University of Maribor, Hoče, Slovenia, (4)CREA, Research Centre for Olive, Citrus and Tree fruit, Forlì, Italy, (5)IMIDA, Instituto Murciano de Investigación y Desarrollo Agrario y Alimentario, Murcia, Spain

In perennial fruit species, phenology and fruit quality traits are highly dependent on environmental conditions. In the context of climate change, a rise of fall and winter temperatures leads to a slowing down of the dormancy release which may result in a dramatic decrease of the production. Also, a rise of spring temperatures induces a significant advance of blooming date, increasing risks of damages by the frost. Moreover, the higher frequency of spring rainy periods has serious consequences for fruit quality, particularly on fruit cracking.

The objective of our study is to identify the genetic determinism of the cherry tree response to the climate change. Several F₁ sweet cherry progenies and a germplasm collection were evaluated for phenological traits and fruit quality. One of the mapping progenies, Regina × Lapins, was evaluated in five sites in order to study the genotype × environment interactions. The progenies were genotyped with the SNP RosBREED arrays and the collection using a genotyping-by-sequencing (GBS) approach. QTL detection and GWAS were performed allowing the identification of the genomic regions associated with the phenological and the fruit quality traits. The comparison of the genomic regions identified by the two analyses was achieved.

For both datasets, we revealed that a genomic region at the end of Chromosome 1 controls both flowering and chilling requirements in sweet cherry, consistent with the observed strong phenotypical correlation between these traits. For the first time, marker-trait associations on several chromosomes were found for traits related to sweet cherry fruit quality.

PO0665: Fruit Species

Genetic Characterization of Blood-Flesh Peach (*Prunus persica* L.) from Chile and Comparative Study of Antioxidants and Micro/Macro Elements Composition during Ripening

Herman Silva, Lab Gen Func & Bioinfo - Fac Ciencias Agro - U de Chile, Santiago, Chile, Basilio Carrasco, Pontificia Universidad Católica de Chile, Santiago, Chile, Rodrigo Corral, Universidad de Chile, Santiago, Chile and Lorena Marchant, Botania SpA, Santiago, Chile

The existence and intensity of the red color in the epicarp and mesocarp in peach fruit is an attractive character due to anthocyanins. These correspond to antioxidants called bioactive compounds responsible for the pigmentation that

covers the color range from red to blue, in leaves, flowers, fruits and roots, and whose benefits are of great potential for human health. In Chile, can be found a population of red mesocarp peaches, located from the Maule Region to the south, whose local name is “durazno betarraga” (blood-flesh peach). The aim of this study was to determine the level of variability and genetic similarity of the peach lines studied, as well as to characterize the phytochemicals and antioxidant compounds present in the fruit. A comparative analysis was carried out, between epicarp and mesocarp of the peach fruits, in two different seasons, corresponding to the 2016-2017 and 2017-2018 season. Fruit quality parameters were evaluated at harvest (weight, size, firmness and soluble solids) and concentrations of total polyphenols, anthocyanins, carotenoids, as well as macro (P, K, Ca, Mg) and micronutrients (Fe, Zn, Mn, B, Cu). These analyzes showed that the blood-peach has high concentrations of anthocyanins (cyanidin-3-glucoside (C3G) and cyanidin-3-rutinoside, C3R) reaching high levels in mesocarp (6201.6 ± 2545.5 mg of C3G eqC3R / kg FW) and epicarp (1073.9 ± 333.1 mg of C3G eqC3R / kg FW) in the first season. Similar results were presented in the 2018 season, being higher in both cases, compared to the varieties of white and yellow mesocarp used as control. For the micro and macronutrients, the analyses did not show significant differences in relation to the control fruits. In terms of the genetic population analysis, a variability genetic analysis using 11,599 SNP markers was performed. The trees showed a low genetic variability and a high genetic similarity.

Acknowledgments: CONICYT, FONDECYT regular N°1160600 and PYT-2015-0390.

PE0666: Fruit Species

Identification of Differentially Methylated Regions in the *Prunus persica* Genome Associated with Mealy Flesh Fruit Due to Long-Term Cold Storage

Alonso Espinoza¹, Karin Rothkegel¹, Dayan Sanhueza², Romina Pedreschi³, Reinaldo Campos-Vargas¹ and **Claudio Meneses**⁴, (1)Centro de Biotecnología Vegetal, Universidad Andrés Bello, Santiago, Chile, (2)Universidad Andrés Bello, Centro de Biotecnología Vegetal, Santiago, Chile, (3)Escuela de Agronomía, Facultad de Ciencias Agronómicas y de los Alimentos, Pontificia Universidad Católica de Valparaíso, Calle San Francisco s/n, La Palma, Quillota, Quillota, Chile, (4)Universidad Andrés Bello, Centro de Biotecnología Vegetal, Santiago, Santiago, Chile

Prunus persica is prone to develop flesh mealiness when exposed to long periods of cold storage (CS). Despite the different attempts to explain mealiness, no satisfactory explanation is available. Abiotic stresses like temperature can induce epigenetic modifications, causing changes in gene expression. This work studied the relationship between mealiness and DNA methylation in *P. persica* causing contrasting individuals for this trait. We used juicy and mealy fruits from an F2 population, and collected samples based on their juice content after CS (30 days at 0 °C). We compared the cold effect on the methylation profile, selecting conditions at harvest (E1) and after CS (E3). Twelve bisulfite-treated genomic libraries were constructed and sequenced, obtaining between 27.4 and 43.9 million reads. Trimmed reads that mapped against the reference genome correspond to a 46.8-55.6 %, while the methylation degree varied between 69.1-72.5%, 49.3-58.9%, and 10-16% in the CpG, CHG, and CHH contexts, respectively. Even when the total methylation level between juicy and mealy fruit was similar in E1, we detected a region of 1.2 Mbp that was significantly higher methylated in juicy than in mealy fruit. This region is located at Chr2:5632244-6907308 bp and contains 17 genes classified by molecular function in “ADP binding” and are related to energetic processes. This study suggests that DNA methylation might have an important role in mealiness susceptibility in *P. persica*. This work was supported by FONDECYT 1160584 and FONDAP Center for Genome Regulation 15090007.

PO0667: Fruit Species

Multi-Omics Identification of Flavor and Aroma Genes in Strawberry

Chris Barbey¹, Kevin M. Folta¹, Vance M. Whitaker² and Seonghee Lee², (1)University of Florida, Gainesville, FL, (2)University of Florida, Wimauma, FL

Flavor is an important consumer trait in the commercial strawberry. However, breeding improvement in this area has been slow due to the immense genetic and chemical complexity of flavor in octoploid strawberry. This work describes multi-omics strategies to identify causal biosynthesis genes at the octoploid subgenome level. Fruit volatile metabolomes were derived from over 800 individuals using statistical alignment techniques on non-targeted

GC/MS data. A subset of 212 individuals were selected for array-based genotyping across eight pedigree-connected biparental populations in multiple seasons. Fifty-five fruit transcriptomes were assembled based on the subgenome-scale octoploid genome to identify candidate genes via trait/transcript correlation and expression-QTL co-segregation. Novel fruit volatile QTL were discovered for methyl anthranilate, methyl 2-hexenoate, methyl 2-methylbutyrate, ethyl butanoate, ethyl hexanoate, mesifurane, and various mono- and sesquiterpenes. These terpenes including linalool, 3-carene, β -phellandrene, α -limonene, linalool oxide, nerolidol, α -caryophellene, α -farnesene, and β -farnesene. An abundantly fruit-expressed methyl transferase is located 0.01 Mb (two genes) from the most-correlated QTL marker shared by three separate methyl ester compounds, including the grape-like methyl anthranilate. In a separate QTL specific to methyl anthranilate, an abundantly fruit-expressed anthranilate synthase gene is located 0.16 Mb (24 genes) from the most-correlated marker. For mesifurane, an epistatic interaction was detected between the known causal gene (O-METHYL TRANSFERASE 1) and a novel QTL likely corresponding to a furaneol glucosyltransferase. Strawberry mono- and sesquiterpenes each co-locate to an identical genomic hotspot containing various terpenoid synthesis pathway components, including the known biosynthesis gene NEROLIDOL SYNTHASE 1 (*FanNES1*). Differences in linalool and other monoterpene levels are partially explained by co-segregation with a *FanNES1* eQTL. Additional evidence show likely quantitative effects from other terpenoid-pathway genes in this narrow hotspot.

PE0668: Fruit Species

Genomic Prediction of Hybrid Performance in Strawberry

Mitchell J. Feldmann, Department of Plant Sciences, University of California, Davis, Davis, CA

Genetic gains for yield have been dramatic in strawberry, an outcrossing allo-octoploid, and have played a pivotal role in the expansion of production over the last half century. Despite evidence for strong directional selection for high yielding cultivars, genetic improvement continues to an effective method to advance crop yields in this highly heterozygous species hypothesized to harbor significant genetic load. Surprisingly, this hypothesis has not been tested, and the importance of heterosis has not been investigated in strawberry. Here, in a 14♀16♂ factorial population, we show that the genetics of yield components are primarily additive, and that high parent heterosis is essentially non-existent in this genetically diverse strawberry population, despite being hypothesized to drive genetic gains for yield in strawberry. Broad-sense heritability ranged from 0.64 to 0.87, whereas narrow-sense heritability ranged from 0.54 to 0.73 for yield components. High parent heterosis was non-significant for 95% of the hybrid progeny tested. We found that the yields of individual hybrids could be accurately predicted using genomic BLUP with additive effects alone and accuracies were not improved by incorporating dominance or epistatic effects. General combining ability was substantially more important than specific combining ability, which did not increase the accuracy of hybrid prediction. Our study shows that yield is highly heritable, well predicted by GCA, and an ideal target for genomic selection, especially since multiple-harvest yield phenotyping costs are exorbitant in strawberry.

PO0669: Fruit Species

A Pentaploid-Based Linkage Map of an Octoploid Strawberry Hybrid Reveals Unusual Patterns of Recombination

Thomas M. Davis¹, Yilong Yang², Daniel Frailey¹ and Lise L. Mahoney³, (1)University of New Hampshire, Durham, NH, (2)UNH, Bedminster, NJ, (3)UNH, Durham, NH

A cross between diploid ($2n = 2x = 14$) *Fragaria vesca* model plant ‘Hawaii 4’ and an octoploid ($2n = 8x = 56$) strawberry hybrid was used to generate a segregating pentaploid progeny population of 178 individuals. A linkage map of the octoploid parent was constructed using markers from the IStraw90 strawberry SNP array (Bassil and Davis et al., 2015), and 29 linkage groups were defined. Unusual patterns of recombination were detected in six individuals, and will be described.

PE0670: Fruit Species

Search for Candidate Gene of Phytophthora Crown Rot Resistance, *FaRPe2*, in Octoploid Strawberry (*F. ×ananassa*)

Saket Chandra¹, Upinder Gill¹, **Youngjae Oh¹**, Young-Hee Noh¹, Nagano Soichiro², Sachiko Isobe², Michael A Hardigan³, Steven J. Knapp³, Vance M. Whitaker¹ and Seonghee Lee¹, (1)University of Florida, Wimauma, FL, (2)Kazusa DNA Research Institute, Kisarazu, Chiba, Japan, (3)Department of Plant Sciences, University of California, Davis, Davis, CA

Phytophthora crown rot (PhCR) caused by *Phytophthora cactorum* is one of the most important soil-borne diseases in cultivated strawberry worldwide. In our previous study, a major quantitative trait locus (QTL) controlling the PhCR resistance, *FaRPe2*, is located on chromosome 7-3 of the octoploid strawberry. In the present study, we conducted fine-mapping of *FaRPe2*. The resistance *FaRPe2* locus was into an approximately 350 kb genomic region. From the transcriptome analysis, we identified candidate genes (NBS-LRR gene, Wall-associated receptor-like kinase, and cyclic nucleotide-gated ion channel gene) associated with the *FaRPe2*-mediated resistance. These genes are highly upregulated in response to *P. cactorum*. Furthermore, we identified putative functional polymorphisms for the candidate genes and developed subgenome-specific high-resolution melting (HRM) markers. The HRM markers were tested for their co-segregation with the *FaRPe2*-mediated resistance in diverse breeding populations. In this presentation, it will be discussed how we precisely locate the *FaRPe2* locus and clone candidate genes using advanced genomics resources.

PO0671: Fruit Species

Dissection of Key Genes Controlling Important Agricultural Traits By Using the Diploid Strawberry *Fragaria vesca*

Chunying Kang, Huazhong Agricultural University, Wuhan, China

The wild diploid strawberry *Fragaria vesca* is an excellent model system for the *Rosaceae* family and the fruit development study. My research interest is to dissect the molecular mechanisms regulating flower and fruit development in strawberry. To this end, we have built a comprehensive transcriptome dataset generated from floral and fruit tissues, identified long non-coding RNAs and alternatively spliced genes in *F. vesca*. Making use of these data, we polished the gene annotation of the *F. vesca* genome. With these genomic and transcriptomic resources, our lab has been using EMS mutagenesis to make a mutant population of *F. vesca*, screening mutants on important agricultural traits, cloning the causative mutations, and investigating gene functions. Through this strategy, we have successfully identified more than 10 genes regulating flower development and fruit quality. One example is *RAP* that is responsible for the foliage and fruit coloration in strawberry. *RAP* encodes a *glutathione S-transferase* (GST) gene that mediates anthocyanin transportation. Among all the homologs in strawberry, *RAP* is most abundantly expressed in the ripening fruit. Transient expression assay demonstrated that *RAP* is the principal transporter of anthocyanins among the paralogs. Moreover, stable over-expression of *RAP* driven by the 35S constitutive promoter in *rap* not only restores anthocyanin accumulation in leaf petiole, but also results in strong coloration in fruit receptacle starting from early developmental stages independent of *FveMYB10*. In addition, knock-out of *RAP* by CRISPR/Cas9 resulted in no leaf petiole coloration in cultivated strawberry, being a promising tool for fruit color breeding. In summary, all the toolkits are available to identify new genes in *F. vesca*, genetically manipulate the homologous genes in cultivated strawberry, and finally create new varieties potentially used for breeding.

PE0672: Fruit Species

Mapping the Primocane-Fruiting Locus in Blackberry (*Rubus* subgenus *Rubus*)

Mitchell E. Armour, University of Arkansas Department of Horticulture, Fayetteville, AR, Margaret Worthington, University of Arkansas, Fayetteville, AR and Bode Adebawale Olukolu, University of Tennessee, Knoxville, TN

While most blackberries (*Rubus* subgenus *Rubus*) are perennial plants that fruit only on second-year canes (floricanes), primocane-fruited cultivars produce fruit on first-year canes. The primocane-fruited trait has allowed for an extended growing season and a wider production range for domestic blackberries. Recent developments in polyploid mapping software have opened up the opportunity to accurately pinpoint the location of this horticulturally important trait within the tetraploid blackberry genome. The primocane-fruited locus was previously mapped to a 12.5 Mbp region that aligns to *R. occidentalis* chromosome 2, but a more precise position could not be determined due to insufficient marker density. The objective of this study is to create a saturated linkage map to further finely map the primocane-fruited locus. A biparental mapping population of 245 individuals was developed

from a cross between a florican-fruited selection with one primocane-fruited parent, and a primocane-fruited breeding selection. The progeny were phenotyped during the 2018 and 2019 growing seasons at the University of Arkansas Fruit Research Station in Clarksville, AR. Fifty of the 245 progeny were scored as primocane-fruited, which fits the 5:1 ratio of florican-fruited to primocane-fruited progeny expected for a nulliplex x duplex cross. The population and parents were subjected by genotyping-by-sequencing to create 2309 filtered SNP markers. The linkage map will be created using PolymapR.

PO0673: Fruit Species

Development of Transcriptome-Derived SSR Markers in Blueberry

Krishnanand Kulkarni¹, Jodi Callwood¹, Lungowe Mulozi¹, Byron Manzanero¹, Nicholi Vorsa^{2,3}, Massimo Iorizzo⁴, Umesh K. Reddy⁵, Sathyanarayana Elavarthi¹ and Kalpalatha Melmaiee¹, (1)Delaware State University, Dover, DE, (2)Rutgers University, Chatsworth, NJ, (3)Department of Plant Biology, Rutgers University, New Brunswick, NJ, (4)Plants for Human Health Institute, Department of Horticultural Science, North Carolina State University, Kannapolis, NC, Madison, WI, (5)Department of Biology, West Virginia State University, Institute, WV

Blueberries (*Vaccinium* spp.) are small fruit crops native to North America but grown commercially in several countries. There has been a steady increase in demand for blueberries, owing to their health benefits attributable to the high content of antioxidant compounds. Despite its importance, limited resources are available for genetic research in blueberry. Simple sequence repeat (SSR) markers from genic regions are useful in genetic analyses as well as provide valuable information on functional sequences and in identifying adaptive genetic variations. In this study, we developed SSR markers from the transcriptome data of two divergent blueberry species, *V. corymbosum* (VC), and *V. darrowii* (VD). More than 135 million high-quality sequence reads (>64.9 million from VC, and >70.3 million from VD) were obtained using Illumina NextSeq500 system. The sequences were *de novo* assembled into 99093 unigenes in VC, and 109193 unigenes in VD after clustering, and were examined for SSR repeats using the MISA tool with default SSR motif criteria. A total of 66549 SSRs in VC and 85051 SSRs in VD were identified. The dinucleotide repeat motifs were more frequent than the other repeat types. A total of 15414 primers for VC and 17765 primers for VD were designed using Primer 3 software, and genomic coordinates for these markers were mapped using recent genome assembly (unpublished). Thirty-six primers have been selected to screen a set of 54 wild and cultivated blueberry genotypes. The markers developed in this study will be a valuable resource for mapping and genetic diversity analyses in blueberry.

PE0674: Fruit Species

Detecting Active Members of the Blueberry Rhizosphere Microbiome Via SIP and rRNA Operon Profiling With the Oxford Nanopore Minion

Seda Mirzoyan¹, Joseph Kawash², **James J. Polashock**³ and Lee Kerkhof¹, (1)Department of Marine and Coastal Sciences/Rutgers University, New Brunswick, NJ, (2)ORISE Scholar/USDA-ARS, Chatsworth, NJ, (3)USDA-ARS, Chatsworth, NJ

Highbush Blueberry (*Vaccinium corymbosum* L.) is a long-lived woody perennial that grows in unique (acidic, sandy, nutrient poor) soils. Although blueberry fields can remain productive for decades, older fields decline in overall plant health and productivity. To begin addressing soil health, we characterized the active members of the rhizosphere microbiome by stable isotope probing (SIP) and rRNA operon profiling. Over 24M raw reads were obtained for eukaryotic/bacterial rRNA operons, which were basecalled with Guppy 3.2.2., and screened by Discontinuous MegaBlast against the Unite ITS and EZBioCloud 16S rRNA gene databases. Preliminary results indicate 1) DNA synthesis from 13C-15N Bioexpress (mostly amino acids) is dominated by 3 eukaryotic fungi-*Banania ogasawarensis*, *Acidea extrema*, and *Acidomyces* sp.; and 2) Declining soil demonstrated 2 to 400—fold more active *Mucor moelleri*, *Mortierella elongata*, *Mortierella rishikeshi*, and *Mortierella chlamydospora* than forest soil or “good” farm soil. 3) The active bacteria were mainly Proteobacteria, Actinobacteria, Firmicutes and only slight differences between “good” and “declining” soil could be observed in the active Proteobacteria and Actinobacteria. 4) However, 9 out of 112 *Bacillus* species exhibited a 2-90 fold increase in abundance between “declining” and “good” soil. These enriched *Bacillus* sp. in “declining” soils included: *B. niacini*, *B. cucumis*, *B. fumarioli*, *B. soli*, *B. novalis*, *B. bataviensis*, *B. pocheonensis*, and *B. drentensis*. This high-resolution approach of

determining active eukaryotes and bacteria in soils may provide a means to define the rhizosphere microorganisms which can impact blueberry plant health.

PO0675: Fruit Species

Vaccinium CAP, a Community-Based Project to Develop Advanced Genetic Tools to Improve Fruit Quality in Blueberry and Cranberry

Massimo Iorizzo, Plants for Human Health Institute, Department of Horticultural Science, North Carolina State University, Kannapolis, NC, Madison, WI, Mary Ann Lila, Plants for Human Health Institute, Kannapolis, NC, Penelope Perkins-Veazie, Professor, Plants for Human Health Institute, North Carolina State University, Kannapolis, NC, Marti Pottorff, Plants for Human Health Institute, North Carolina State University, Chad Finn, USDA-ARS, Horticultural Crops Research Unit, Corvallis, OR, Nicholi Vorsa, Rutgers University, New Brunswick, NJ, Patrick Edger, Department of Horticulture, Michigan State University, East Lansing, MI, Nahla Bassil, USDA-ARS National Clonal Germplasm Repository, Corvallis, OR, Patricio R. Munoz, Horticultural Sciences Department, University of Florida, Gainesville, FL, Juan Zalapa, University of Wisconsin-Madison, Madison, WI, Rosa Karina Gallardo, School of Economic Sciences, Puyallup Research and Extension Center, Washington State University, St. Puyallup, WA, Amaya Atucha, Department of Horticulture, University of Wisconsin-Madison, Madison, WI, Dorrie Main, Department of Horticulture, Washington State University, Pullman, WA, Lara Giongo, Fondazione Edmund Mach, San Michele all'Adige, Italy, Changying Li, College of Engineering, University of Georgia, Athens, GA, James J. Polashock, USDA-ARS, Chatsworth, NJ, Charles Sims, Department of Food Science and Nutrition, Aquatic Food Production Pilot Plant, University of Florida,, Elizabeth Canales, Mississippi State University, Lisa DeVetter, Washington State University, Pullman, WA, David Chagné, The New Zealand Institute for Plant & Food Research Limited, Palmerston North, New Zealand, Richard Espley, The New Zealand Institute for Plant & Food Research Limited, Auckland, New Zealand, Michael Coe, Cedar Lake Research Group and PAG Authors'Group

Vaccinium crops (primarily blueberry and cranberry) provide vital contributions to the U.S. economy and numerous social and health benefits. Although production and consumption of these crops continue to expand worldwide, U.S. *Vaccinium* industries face numerous challenges to maintain profitability. Stakeholders have asserted that breeding cultivars with improved fruit quality is a high priority for continued success. These traits include fruit firmness, flavor, shelf life, and appearance. *Vaccinium* breeders routinely select for these traits, however, they have little empirical data to assign a level of importance to specific fruit characteristics (FC) relative to consumer preferences, decay or deterioration during production, processing and distribution. Breeders also possess limited tools to select for a higher quality fruit. The VacciniumCAP project was funded by USDA-NIFA-SCRI to create a nationwide coordinated transdisciplinary research approach to develop marker-assisted selection capacity in *Vaccinium* breeding programs, and to select for and pyramid FCs which enhance fruit quality and market value. The project objectives are to: 1) Establish genomic resources to enable effective association mapping studies in blueberry and cranberry; 2) Discover DNA markers and fruit characteristics that maximize industry profitability and match consumer preferences in blueberry and cranberry; 3) Deliver molecular and genetic resources to improve blueberry and cranberry fruit quality traits that maximize industry profitability and match consumer preferences; 4) Assess the potential socio-economic impact of blueberry and cranberry fruit quality improvements on market demand; and 5) Engage U.S. *Vaccinium* stakeholder groups to transfer advanced phenomic and genomic tools to build a more efficient cultivar development system.

PE0676: Fruit Species

Manipulation of Flowering Pathway Genes for High Fruit Productivity

Guo-Qing Song, Michigan State University, East Lansing, MI

Blueberries contain high amounts of antioxidants known to be important for human health. Developing new cultivars with different chilling requirement, high cold/heat tolerance, and for high yield are among the top priorities in blueberry breeding due particularly to the anticipation of climate changes and the rapidly expanding market need of blueberry products. Overexpression of a blueberry *DWARF AND DELAYED FLOWERING 1* (*VcDDF1*)

increased freezing tolerance without a trade-off impact on yield. Turning on a blueberry a *RESPONSE REGULATOR 2*-like gene (*VcRR2*) in a mutant caused by the *VcDDF1* transgene insertion resulted in reduced chilling requirement for flowering and a high yield potential. The mutant provides an outstanding material to study chilling-mediated flowering mechanism in woody plants. Transgenic blueberries overexpressing a blueberry *FLOWERING LOCUS T* (*VcFT*) facilitate FAST-TRACK blueberry breeding through grafting. Grafting on *VcFT*-overexpressing blueberry plants promoted floral bud formation in nontransgenic scions and thus not only demonstrates that hormones are involved in *FT*-induced long-distance transport of the florigenic signals but also provides a new approach to increase blueberry yield. Overexpression of the K-domain of a blueberry *SUPPRESSOR OF OVEREXPRESSION OF CONSTANS 1* gene increased berry productivity through the interaction of MIKCMADS-box genes. This new K-domain technology is being tested in corn and works as well. Overall, we demonstrate that manipulation of flowering pathway gene(s) or hormone synthesis pathway gene(s) is a powerful approach to increase fruit/crop productivity.

PO0677: Fruit Species

Towards Understanding of Genome Evolution of Southern Highbush Blueberry By ddRAD-Seq

Soichiro Nishiyama¹, Mao Fujikawa¹, Hisayo Yamane¹, Kenta Shirasawa², Ebrahiem Babiker³ and Ryutaro Tao¹, (1)Graduate School of Agriculture, Kyoto University, Kyoto, Japan, (2)Kazusa DNA Research Institute, Kisarazu, Japan, (3)USDA-Agricultural Research Service, Poplarville, MS

Southern highbush blueberry (SHB) is a blueberry cultivar group developed by crosses between northern highbush [NHB (*V. corymbosum* L. 2n=4x=48)] and indigenous blueberry species in Florida to expand the geographic limits of highbush blueberry production. Due to the multiple use of interspecific hybridization in breeding process, most SHB cultivars assumed to contain genomic segments introduced from one or more of the *Vaccinium* species. The complexity in the polyploid nature of highbush blueberry has hindered full understandings of the genetic diversity and genome evolution of SHB. Here we genotyped 108 accessions of SHB, 19 accessions of NHB and 11 accessions of rabbiteye blueberry (RE) by ddRAD-seq. SNPs were called as dominant marker with 5% threshold of alternative allele in this study, which yielded in 119,072 genome-wide SNP loci with biallelic variants. To evaluate the distance of SNP pairs with substantial association, we analyzed the decrease of the squared correlation coefficient values against physical distance for each SNP allele with all other SNPs in the chromosome, and calculated the maximum distance with the association by applying quantile regression and cubic spline. In the SHB population, long potential associations in pairs of SNPs with distance up to approximately 10 Mbp were found in chromosome 1, 4, 5, 6, 7, 8. Those associations tend to be located on the center of the chromosome.

PE0678: Fruit Species

Evidence of a Reciprocal Heterozygous Translocation Leading to Gamete Abortion in Cranberry

Joseph Kawash, ORISE/USDA ARS, Chatsworth, NJ, James J. Polashock, USDA-ARS, Chatsworth, NJ and Nicholi Vorsa, Rutgers University, New Brunswick, NJ

Pollen tetrad analysis of a fruit rot resistant germplasm accession, US89-3, indicated the presence of heterozygosity for a reciprocal translocation, prompting a marker investigation of the translocated regions. The segregation of markers (SNPs) within a US89-3 x Crimson Queen progeny population were analyzed using genotyping-by-sequencing. Contrasting the genetic maps of families derived from US89-3 to those lacking the US89-3 parental background revealed that US89-3 contained a genetic reorganization between LG5 and LG6. In cranberry, the four resulting gametes (pollen), of microgametogenesis are shed as a tetrad. Tetrad pollen viability analysis revealed a trimodal distribution of 4's, 2's and 0's classes resulting from genetic imbalance when cross-overs occur proximal to the translocation event in individuals heterozygous for the translocation. Whole genome sequencing using Illumina paired end reads and alignment to an updated cranberry reference genome (487Mb, 124 contigs, N50 15Mb) was performed to better locate the translocation breakpoint in US89-3. GROM was used to identify potential translocation events in the NGS data. Four unique reciprocal heterozygous translocations were identified in NGS data between LG5 and LG6. There were no predicted genes identified crossing these translocation breakpoints, however eight gene fragments were found within 1kb in each direction of breakpoints.

PO0679: Fruit Species

citSatdb: Genome Wide Simple Sequence Repeat (SSR) Marker Database of Citrus Species for Germplasm Characterization and Crop Improvement

Naveen Duhan, Manish Mehsram, Cristian D Loaiza and Rakesh Kaundal, Utah State University, Logan, UT

The *Citrus* genus is the world's most widely cultivated fruit crop. Microsatellites or simple sequence repeats (SSRs) are popular co-dominant markers which play an important role in crop improvement. With the goal of enhancing the genomic resources in *citrus*, we identified SSRs in genomes of 6 *citrus* species and characterized their frequency and distribution in different genomic regions. The advent of next generation sequencing technologies and advancement in computational approaches have made possible the discovery of markers in bulk. The potential polymorphic SSR markers identified by cross species transferability could be used for genetic diversity and population distinction in other species. We describe an online database, citSatdb (<http://bioinfo.usu.edu/citSATdb/>) having the highest number (~17,23,036) of putative SSR markers from *Citrus* genus so far, represented by six species: *Citrus sinensis*, *Citrus clementina*, *Citrus maxima*, *Citrus medica*, *Citrus ichangensis* and *Atlantia buxifolia*. The database is based on a three tier approach using MySQL, PHP and Apache. The markers can be searched using multiple search parameters including chromosome/scaffold number(s), motif types, repeat nucleotides(1-6), SSR length, pattern of repeat motif and chromosome/scaffold location. Cross species transferability of selected markers can be checked using e-PCR. Further markers can visualized using Jbrowse. These markers can be used for Distinctness, Uniformity, and Stability (DUS) tests of variety identification, marker assisted selection (MAS), gene discovery, QTL mapping, and germplasm characterization. The database represents a source of markers for developing and implementing new approaches for molecular breeding, which are required to enhance *Citrus* productivity.

PE0680: Fruit Species

A Whole Genome Association Study in Mandarin Hybrids

Javier Terol, Carles Borredá, Estela Pérez-Román, Daniel Gonzalez-Ibeas, Victoria Ibañez, Francisco R Tadeo and Manuel Talon, Centro de Genómica, Instituto Valenciano de Investigaciones Agrarias, IVIA, Moncada, Valencia, Spain

A segregating population of 667 hybrids from a cross between Clemenules x Nadorcott has been established in three different locations in Spain. The population has been genotyped by sequencing (GBS) and phenotyping is in its second year. The data will be used to obtain markers useful for marker assisted breeding.

This phenotyping protocol will focus in the main traits related to vegetative and reproductive stages, with special attention to fruit quality and productivity, as well as efficiency in the use of water, resistance to heat waves or the presence of symptoms of biotic and abiotic stresses.

We will use these data to perform genome wide association studies, that will include the following steps: definition of phenotypes (quantitative vs. presence/absence), filtering of samples with Low Call Rate, studies of stratification of the population (Structure), quality of the SNPs (call rate, frequency of the minor allele), statistical analysis

(Additive Model, Mixed Linear Model, Multi-Locus Model Analysis), correlation studies, Manhattan plots (P value), selection of significant SNPs, and study of genomic regions and nearby genes.

The GBS has produced a total of 11103 SNPs useful for GWAS and some phenotyping data are already available. Some preliminary results will be shown.

PO0681: Fruit Species

A New Satsuma Mandarin Cultivar ‘Jedae’

In-Jung Kim, Faculty of Biotechnology, Jeju National University, Jeju-si, Jeju Special Self-Governing Province, South Korea

This research relates to a Jedae satsuma mandarin, which is a novel variety of a mutant satsuma mandarin citrus. A novel mutant citrus plant was bred by emitting gamma rays at a bud on a branch of an early-ripening satsuma mandarin of *Citrus unshiu* Marc. cv., Miyagawa-wase, grafting the same and subjecting the same to asexual reproduction, resulting in a very close resemblance with the *C. unshiu* Marc. cv. Miyagawa-wase with respect to flowering, bearing fruit, a sweetness/acidity value, and the like, but the pericarp of the fruit is not smooth and the amount of flavonoid ingredients is changed. That is, the present work relates to ‘Jedae mandarin’, a new citrus variety of early-ripening satsuma mandarin with altered fruit shape and flavonoid content. Thus our developed new citrus cultivar can be used as a novel variety having high economic value.

PE0682: Fruit Species

Development of Molecular Markers to Discriminate New Citrus Cultivars using RAPD Analysis

MyungUn Jung, Department of Biomaterials Engineering, Graduate School, Jeju National University, Jeju, South Korea and **In-Jung Kim**, Faculty of Biotechnology, Jeju National University, Jeju-si, Jeju Special Self-Governing Province, South Korea

We have developed several citrus mutant cultivars through artificial mutation. The RAPD method is one of the effective methods to find genetic diversity or discrimination methods among cultivar. It can be used to analyze the terminal base sequence of the amplified region and then to reproduce it by using a specific primer, SCARs (sequence characterized amplified). New citrus varieties used in this study are our developed mutants (Ara mandarin, Jedae mandarin) induced from Satsuma Mandarin (*Citrus unshiu* Marc. cv. Miyagawa-Wase early) in our laboratory. As a result, PCR method using RAPD primer was used to identify the control group Marc. cv. Miyagawa-Wase early. We used the 44 primer combinations for Jedae mandarin, and 48 primer combinations for Ara mandarin to compare with control group. We have observed specific bands through PCR. In order to convert the RAPD markers into PCR-based SCAR markers using them as a primer combination for SCAR marker development, specific bands amplified in the gel were cloned, and sequencing analysis was performed to obtain nucleotide sequence information. Through the repetitive PCR process, we have confirmed the availability of putative SCAR marker for cultivar discrimination.

PO0683: Fruit Species

Comparative Transcriptomes and Metabolomes Reveals Genetic Regulation of Low Furanocoumarin in Developing Fruits of UF 914 and Ruby Red Grapefruit

Qibin Yu, Zhibin Wang, Matthew R Mattia and Fred G. Gmitter, University of Florida, IFAS-CREC, Lake Alfred, FL

Commercial grapefruits have furanocoumarin compounds which could interact with a wide range of medications. This concern has a substantial negative impact on grapefruit marketing. The grapefruit hybrid cultivar UF 914 has significantly lower furanocoumarin content than commercial grapefruits, which could significantly reduce its medication interaction. Despite the importance of coumarins for plant defense against insects and human uses, major details of their biosynthesis have remained unresolved. Previous studies indicated that multiple CYP450 enzymes were involved in furanocoumarin synthesis. In this study, fruits of these varieties at 62 days post anthesis (DPA; Stage I cell division), at 171 DPA (Stage II, cell expansion), and 286 DPA (Stage III, mature) were collected.

Transcriptomes of the three stages were profiled using RNA sequencing, and furanocoumarins were also measured. Major production of furanocoumarins was significantly lower in UF 914 than Ruby Red grapefruit at all three stages. Functional analysis of the differentially expressed genes (DEGs) indicated that genes involved in the phenylpropanoid pathway were highly upregulated in Ruby Red, but not in UF 914. A CYP450 gene associated with furanocoumarin production was identified. The function of the CYP450 gene in biosynthesis of furanocoumarins was studied using transgenic plants and transient expression in fruits of UF 914 and Ruby Red grapefruit.

PE0684: Fruit Species

ERF109 of *Poncirus trifoliata* (L.) Raf. Functions in Cold Tolerance By Regulating Prx1 Involved in Antioxidative Process

Ji-Hong Liu, Huazhong Agricultural University, Wuhan, China

Ethylene-responsive factors (ERFs) have been revealed to play essential roles in a variety of physiological and biological processes in higher plants. However, functions and regulatory pathways of most ERFs in cold stress remain largely unclear. Here, we identified *PtrERF109* of trifoliolate orange (*Poncirus trifoliata* (L.) Raf.) and deciphered its role in cold tolerance. *PtrERF109* was drastically up-regulated by cold, ethylene and dehydration, but repressed by salt. *PtrERF109* was localized in the nucleus and displayed transcriptional activity, and the C terminus is required for the activation. Overexpression of *PtrERF109* conferred enhanced cold tolerance in transgenic tobacco and lemon plants, whereas VIGS (virus-induced gene silencing)-mediated suppression of *PtrERF109* in trifoliolate orange led to increased cold susceptibility. *PtrERF109* overexpression caused extensive transcriptional reprogramming of several suites of stress-responsive genes. *Prx1* encoding class III peroxidase (POD) was one of the antioxidant genes exhibiting the greatest induction. *PtrERF109* was shown to directly bind to the promoter of *PtrPrx1* (trifoliolate orange *Prx1* homologue) and positively activated its expression. In addition, the *PtrERF109*-overexpressing plants exhibited significantly higher POD activity and accumulated dramatically less H₂O₂ and were more tolerant to oxidative stress, whereas the VIGS plants exhibited opposite trends, in comparison with wild type. Taken together, these results indicate that *PtrERF109* as a positive regulator contributes to imparting cold tolerance by, at least partly, directly regulating the POD-encoding gene to maintain a robust antioxidant capacity for effectively scavenging the ROS. Our findings gain insight into better understanding of transcriptional regulation of antioxidant genes in response to cold stress.

PO0685: Fruit Species

Greeningdb: Database of Protein Features and Protein-Protein Interactions of the Bacteria Causing HLB

Cristian D Loaiza, Naveen Duhan and Rakesh Kaundal, Utah State University, Logan, UT

The *Citrus* genus comprises some of the most important and commonly cultivated fruit plants. Within the last decade, citrus greening disease (also known as Huanglongbing or HLB) has emerged as the biggest threat to the citrus industry. This disease does not have a cure yet, therefore many efforts have been made as an attempt to find a solution to this devastating condition. There are challenges in the generation of high-yield resistant cultivars, in part due to the limited and sparse knowledge about the mechanisms that are used by the *Liberibacter* bacteria to proliferate the infection in *Citrus* plants.

Here we present GreeningDB (<http://bioinfo.usu.edu/GreeningDB/>). GreeningDB is a database implemented to provide the annotation of the proteins from several of the *Liberibacter* strains, as well as the host-pathogen "comparactomics" tool, a platform to compare the predicted interactomes of HLB host-pathogen systems. This database has been built to deliver a user-friendly interface, network visualization and links to other related resources. We hope that by providing these functionalities, GreeningDB can become a central resource to retrieve HLB protein information, and thus could aid to the community pursuing into the development of breeding and/or molecular-based strategies to mitigate this disease's impact.

PE0686: Fruit Species

New Strategy to Evaluate Putative Susceptibility Genes to Citrus Canker Induced By PthA4

Reinaldo Rodrigues de Souza-Neto¹, Fernanda Nogales da Costa Vasconcelos², Doron Teper², Tirtha Lamichhane², Celso Eduardo Benedetti³, Marco Aurélio Takita⁴, Nian Wang⁵ and Alessandra Alves De Souza⁶, (1)Universidade Estadual de Campinas/ Centro APTA Citros Sylvio Moreira - Brazil, Cordeirópolis, Brazil, (2)University of Florida, Lake Alfred, FL, (3)Centro Nacional de Pesquisas em Energia e Materiais, Campinas, Brazil, (4)Citrus Center, IAC, Sao Paulo State Secretary of Agriculture and Supply, Cordeirópolis, SP, Brazil, (5)University of Florida, IFAS-CREC, Lake Alfred, FL, (6)Centro APTA Citros Sylvio Moreira, CORDEIROPOLIS, Brazil

The pathogenicity of *Xanthomonas citri*, causal agent of citrus canker, is associated with the secretion of effectors. One class of these effectors is Transcription Activator-Like Effectors (TALEs). These effectors bind to specific region (named Effector Binding Element (EBE)) of the promoter of target genes. Genes induced by these effectors, whose induction promotes disease development, are called susceptibility genes. *LATERAL ORGAN BOUNDARIES 1 (LOB1)* has been identified to be a canker susceptibility gene induced by PthA4, the predominant TALE responsible for canker symptoms. We hypothesize that PthA4 targets additional susceptibility genes besides *LOB1* to facilitate its infection. Here, we aimed to develop a new strategy for identification of additional canker susceptibility genes induced by PthA4. The original and a modified promoter of *LOB1* (EBE for PthA4 replaced by predicted but not experimentally validated EBE for PthA3 (LOB1M)) were used to drive the expression of GUS in co-infiltration experiment with effectors PthA4 and PthA3. High GUS activity was observed when co-infiltrated PthA4 with *LOB1* promoter. As negative control PthA3 co-infiltrated with *LOB1* promoter did not show GUS activity, because there is not EBE region for PthA3 in the original promoter. When we used modified promoter, which contains EBE for PthA3 and co-infiltrated with PthA3, higher GUS activity was observed compared to negative control infiltrated with PthA4, because there is not EBE for PthA4 in modified promoter. Thus, we developed a new method for identification of putative susceptibility genes induced by PthA4. In addition, we showed that predicted EBE could be recognized by PthA3 and driving transcription in non-original sequences.

PO0687: Fruit Species

Response of *Citrus sinensis* to *Diaphorina citri* Inoculation of *Candidatus Liberibacter asiaticus*: A Systems Biology Perspective

Rachel L. Lombardi¹, John S. Ramsey^{2,3}, Jaclyn Mahoney^{2,3}, Michelle Heck^{2,4} and Carolyn M. Slupsky^{1,5}, (1)Department of Food Science, University of California, Davis, Davis, CA, (2)USDA ARS, Ithaca, NY, (3)Boyce Thompson Institute for Plant Research, Ithaca, NY, (4)Cornell University, Ithaca, NY, (5)Department of Nutrition, University of California, Davis, Davis, CA

Huanglongbing (HLB) is a fatal citrus disease currently threatening all commercially relevant citrus varieties worldwide. In the United States, the putative causative agent, *Candidatus Liberibacter asiaticus* (CLas), is transmitted by the vector *Diaphorina citri*, also known as the Asian Citrus Psyllid (ACP). Progress towards understanding CLas and its role in HLB has been limited due to the inability to culture the bacterium, the inability to confidently detect it in diseased trees, and the variation in citrus host response. In this study, we utilized metabolomics, transcriptomics, and proteomics to compare the citrus response to psyllid feeding with and without concurrent CLas infection. This study will be one of the first longitudinal investigations to utilize a complete multi-omics platform to investigate the response of sweet orange (*Citrus sinensis* L. Osbeck) to inoculation of CLas using its natural psyllid vector. A comprehensive approach to understanding the early plant response to ACP transmitted CLas infection will aid development of early detection technologies and inform on ways to develop resistant cultivars.

PE0688: Fruit Species

The Genetic Control of Long Shelf Life in Melon (*Cucumis melo* L.)

Shumpei Goda¹, Toshiki Hatanaka¹, Takashi Akagi², Yasutaka Kubo², Ryohei Nakano³ and Koichiro Ushijima², (1)Okayama university, Okayama, Japan, (2)Okayama University, Okayama, Japan, (3)Kyoto University, Kyoto, Japan

Melons (*Cucumis melo* L.) are generally classified into climacteric type fruit, in which that ethylene trigger and promote the ripening, and have poor shelf life because of rapid flesh softening. The features of fruits ripening are

highly diverged among melon cultivars, and some cultivars/strains have long shelf term and have been used as the genetic resources for breeding. B2, Earls net melons, have long shelf life, although the autocatalytic ethylene was evolved during ripening. To investigate the mode of inheritance of long shelf life in B2, we prepared the segregated population by crossing B2 (female) and Harukei-3 (male), a rapid softening cultivar, and the flesh softening rate was detected by a nondestructive acoustic vibration method. All F1 hybrids from the cross, had long shelf life. The softening rate are comparable to that of B2, the seed parent. The F2 population could be divided into slow and rapid softening type fruits at the ratio of 3:1. This ratio suggested a possibility that the long shelf life of B2 is dominant trait and controlled by a single diallelic locus. Indeed, investigation of F3 population demonstrated that some homozygous lines were obtained in F2. The simple inheritance of the shelf life trait of B2 is expected to be attractive for genetic resource, but we analyzed small population derived from only one cross in this study. Large population and other cross combination will be helpful for verifying the hypothesis.

PO0689: Fruit Species

Selection of Housekeeping Genes for qRT-PCR Studies During Compatible Heterograft Union Formation in Watermelon/Squash Graft Combination

Hui Song, Ningbo Academy of Agricultural Sciences, Ningbo, China

Quantitative real-time PCR has become standard method for gene expression analysis. However, the expression profile of a target gene may be misinterpreted because of unstable expression of the reference genes under different experimental conditions. Thus, a systematic evaluation of reference genes is necessary before experiments are performed. In the present report we describe the first systematic evaluation of potential housekeeping genes during heterograft union formation in watermelon/squash graft combination to identify which are the most reliable for transcript quantification by qRT-PCR. In this study, 14 putative reference genes reported in cucumber, melon, squash and loofah were chosen for identifying expression stability using geNorm, Norm-Finder, and BestKeeper statistical algorithms. A set of 11 samples representing original tissues, callus and graft union of watermelon and squash at different days after grafting were prepared. First, three biological duplications of the 11 samples' RNA and cDNA were isolated and prepared for three tubes. Second, the qRT-PCR amplification efficiency was analyzed by cDNA template of 20 days watermelon callus for candidate reference genes. Then, expression data from each reference gene were evaluated with three complementary approaches based on different statistical procedures. Our analysis suggests that ACT and TUA genes exhibited the most stable expression across all of the tested samples, which could be used along or combined for study on the expression of differential gene during compatible heterograft union formation in watermelon/squash graft combination.

PE0690: Fruit Species

Construction of Reference Genome Sequences of Cultivated-Type 'SBA' and Citron-Type PI189225 Accession in Watermelon

Yoon Jeong Jang¹, Minseok Seo², Bal-Kum Han¹ and Gung-Pyo Lee¹, (1)Dept of Plant Science and Technology, Chung-Ang University, Ansung, South Korea, (2)Dept. of Computer Convergence Software, Korea University, Sejong Metropolitan, South Korea

Watermelon (*Citrullus lanatus*) has been known to abundant nutrition resource of citrulline, phytochemical lycopene and essential amino acid for human health. Various breeding studies based on genetics and transcriptomics approaches in watermelon have been conducted, and more accurate reference genome and gene annotation will be required for more accurate research. Previously established reference genomes and gene annotations for two watermelon cultivars were constructed based on short-read sequencing technique. However, there are clear limitations in the identification of precise genome and isoform structure due to the technical limitation derived from short-read assembly. Also, the importance of wild species genome information has been noticed to breed introgressed lines for improving cultivars. Here, we present high-quality reference sequences of 'SBA' (*C. lanatus* ssp. *vulgaris*) and PI189225 (*C. amarus*). 'SBA' is an elite line for producing commercial F1 hybrid in Korea. Citron-type PI189225, one of useful wild watermelons, has been mainly used for the generation of pathogen-resistant cultivar breeding, but the full genome sequence has not been reported. Reference genome of two species was respectively constructed using Illumina-Seq and single molecule real-time (SMRT)-Seq of PacBio platforms. Superscaffolds were generated using optical mapping data of BioNano Irys system. The scaffolding order was

refined and confirmed by the recombination map data from the crossing of cultigens and PI18925. Finally, we successfully identified the sequence of full-length transcriptome using PacBio single-molecule long-read isoform sequencing technique for the first time in watermelon. We believe that these resources will accelerate functional genetic studies, provide valuable reference for comparative genomics, evolutionary researches and molecular breeding programs in watermelon.

PO0691: Fruit Species

Novel SNP Markers Associated with Flesh Colors in Watermelon

Geung-Joo Lee and Saminathan Subburaj, Chungnam National University, Daejeon, South Korea

Recently we identified loci that are associated with orange flesh color and higher carotenoid content. They were residing on chromosome 1, 6, 8 and 10, among which the latter three loci have not been reported previously.

PE0692: Fruit Species

Phenotypic Variability and Genome-Wide Association Analysis of Downy Mildew (*Pseudoperonospora cubensis*) Resistance in a Pre-Breeding Watermelon (*Citrullus amarus*) Collection

Dennis N. Katuuramu, USDA - ARS, Charleston, SC

Cultivated sweet watermelon (*Citrullus lanatus*) is an important vegetable crop for millions of people around the world. There are limited sources of resistance to economically important diseases within *C. lanatus*, whereas *Citrullus amarus* has a reservoir of traits that can be exploited to improve *C. lanatus*. We screened 123 *C. amarus* accessions for resistance to downy mildew over two tests (environments). The accessions were genotyped with 1,902,395 single nucleotide polymorphic (SNP) markers. Genome-wide association study approach was deployed to uncover marker-trait associations and candidate genes underlying resistance to downy mildew. Our results indicate the presence of wide phenotypic variability of 1.14 – 57.76 % leaf area infection (50.7-fold variation) for downy mildew resistance in the *C. amarus* diversity panel. Broad-sense heritability estimate was 55 % implying presence of moderate genetic effect for resistance to downy mildew. The significant SNP markers associated with resistance to *P. cubensis* were located on chromosomes 4, 6, and 10. This novel information will be useful in understanding the genetic architecture of the *P. cubensis*-*Citrullus* spp. patho-system as well as development of resources for genomics-assisted breeding for resistance to downy mildew in watermelon.

PO0693: Fruit Species

QTL and Transcriptomic Analyses Implicate Cuticle Transcription Factor SHINE As a Source of Natural Variation for Epidermal Traits in Cucumber Fruit

Stephanie Rett-Cadman¹, Marivi Colle¹, Ben N. Mansfeld², Cornelius S. Barry², Yuhui Wang³, Yiqun Weng⁴, Lei Gao⁵, Zhangjun Fei⁵ and Rebecca Grumet¹, (1)Department of Horticulture, Michigan State University, East Lansing, MI, (2)Michigan State University, East Lansing, MI, (3)Department of Horticulture, University of Wisconsin, Madison, WI, (4)USDA ARS/ University of Wisconsin, Madison, WI, (5)Boyce Thompson Institute, Cornell University, Ithaca, NY

The fruit surface is a unique tissue with multiple roles influencing fruit development, post-harvest storage and quality, and consumer acceptability. Fruit surfaces can vary markedly among species, cultivars, and developmental stage. This study examined developmental changes and natural variation of cucumber (*Cucumis sativus*) fruit surface properties using two lines that differ for these traits and for which draft genomes and a single nucleotide polymorphism (SNP) array are available: Chinese fresh market type, Chinese Long '9930' (CL9930), and pickling type, 'Gy14'. Thin-section samples prepared from the mid-region of fruit were evaluated for cuticle thickness and depth of intercalation between epidermal cells, epidermal cell size and shape, and number and size of lipid droplets. 'Gy14' exhibited columnar shaped epidermal cells, thicker cuticle, deeper cuticular intercalations and greater number and size of lipid droplets. In both lines maximal deposition of cuticle and increase in epidermal size coincided with exponential fruit growth and was largely completed by ~16 days post pollination. Phenotypic analysis and quantitative trait locus mapping (QTL) of fruit from an F_{7:8} Gy14 × CL9930 recombinant inbred line

(RIL) population identified strong QTL for epidermal cell height, cuticle thickness, intercalation depth, and diameter of lipid droplets co-localized on chromosome 1. Fine mapping of an extended RIL population using SSR and KASP markers combined with gene expression profiling suggested a small number of candidate genes. Tissue specificity, developmental analysis of expression, allelic diversity and gene function implicate the regulatory factor *CsSHINE1/WIN1* as a source of natural variation for cucumber fruit epidermal traits.

PE0694: Fruit Species

Transcriptome Analysis of Maternally-Transmitted Cold Tolerance in Cucumber

Madeline W. Olberg, University of Wisconsin - Madison, Madison, WI and Michael J. Havey, USDA-ARS and University of Wisconsin, Madison, WI

Cucumber is a warm-season crop that can be severely damaged by short periods of cold temperatures. Cold tolerant cucumbers would benefit growers by preventing crop loss in inclement weather as well as by allowing for earlier planting and harvest, thus avoiding heavy late-season disease pressure. Maternally-inherited cold tolerance has been reported in the heirloom cucumber cultivar, 'Chipper.' In cucumber, the chloroplast, mitochondrial, and nuclear genomes are maternally, paternally, and biparentally transmitted, respectively, indicating that this cold tolerance may be conditioned by the chloroplast genome. Phenotypic analysis of reciprocal hybrids between doubled haploids (DH) of cold-tolerant 'Chipper' and susceptible 'Straight 8' and 'Marketmore' revealed that cold recovery is maternally transmitted from 'Chipper'. Total nuclear RNA was extracted and sequenced from reciprocal hybrids with identical nuclear genotypes, revealing similar expression profiles for the cold-tolerant hybrids after cold treatment. Further analysis of a non-synonymous SNP in a chloroplast ATPase subunit of 'Chipper' could further elucidate the mechanism of this cold recovery phenotype. Identification of the genetic basis of cold tolerance in 'Chipper' would provide potential targets of selection for cold tolerance in cucumber and other warm-season crops.

PO0695: Fruit Species

CsPhyB and CsGA20ox-2 Coordinate Hypocotyl Elongation in Cucumber

Jianyu Zhao, Department of Horticulture, Madison, WI and Yiqun Weng, USDA ARS/ University of Wisconsin, Madison, WI

In plants, light and gibberellins (GAs) mediate many essential and partially overlapping development processes including hypocotyl elongation and cotyledon opening. We identified two spontaneous mutations, *LONG HYPOCOTYL1* (*Lh1*), and *LONG HYPOCOTYL2* (*Lh2*) in natural populations of cucumber, *Cucumis sativus* L. Both mutations are recessively inherited. Map-based cloning revealed that *Lh1* encodes an ortholog of the Arabidopsis PHYTOCHROME B (CsPhyB), and *Lh2* encodes GA20oxidase-2 (CsGA20ox-2), a class I GA 20-oxidase in the GA biosynthetic pathway. The long hypocotyl in *lh1* and *lh2* was due to a 7bp and a 1bp deletion, respectively, in the first exon of both genes resulting in premature translation and truncated proteins. Both *lh1* and *lh2* mutants exhibited altered responses to treatments by GA and PAC (inhibitor of GA biosynthesis) suggesting the roles of CsphyB in GA-mediated hypocotyl elongation. We found interactions among CsphyB, CsPIFb (a member of phytochrome interacting factor family), and CsDELLA1 (a negative regulator of GA signaling). CsPIF1b from the wild type, but not from the mutants was able to bind to the G-box motif of *CsGA20ox2* promoter to regulate its transcription. A working model was proposed in which CsphyB modulates GA level, thus hypocotyl elongation through regulating *CsGA20ox2* transcription via through CsphyB-CsPIF1b-CsDELLA1 interactions.

PE0696: Fruit Species

Integrating Chemical Ecology and Genomics to Develop Breeding Strategies for Pest Resistance in *Cucurbita pepo*

Lauren Brzozowski¹, Anurag Agrawal², Michael A. Gore¹ and Michael Mazourek¹, (1)Plant Breeding and Genetics Section, School of Integrative Plant Science, Cornell University, Ithaca, NY, (2)Department of Ecology and Evolutionary Biology, Cornell University, Ithaca, NY

Pest-resistant crops are essential in integrated pest management, but breeding progress for resistance is constrained by low heritability and a limited understanding of resistance mechanisms. Here we present integrative research to

breed for resistance to the Cucurbitaceae-specialized *Acalymma vittatum* (striped cucumber beetle) in *Cucurbita pepo* (zucchini, summer squash) that draws on fundamental chemical ecology principles and plant genomics. First, within the Cucurbitaceae, cucurbitacins, bitter terpenoids, are known to attract *A. vittatum*; however, their effect within agricultural systems is unknown. Using HPLC-MS to measure multiple cucurbitacins, we concurrently mapped biosynthetic genes and conducted bidirectional selection. We discovered previously unreported tissue-specific biosynthesis, but demonstrated that lack of pleiotropic effects between economically important tissues (cotyledons, leaves) renders cucurbitacins only a limited target for plant breeders. This provides impetus for further mechanistic experiments to identify additional breeding targets. Thus, over three years of chemical ecology field experiments and complementary untargeted metabolomics (GC-MS), we indeed identified a new foliar volatile that acts as a pest deterrent. Publicly available genomic resources (sequences, core collections) will allow for selective phenotyping of this trait, and rapid introgression into breeding populations. However, quantifying insect resistance is resource intensive, making genomic selection an attractive alternative to the previously described mechanistic approaches. We used genomic selection for the low heritability trait of field-measured *A. vittatum* resistance, and achieved moderate cross-fold validation accuracy, and empirical gains in the field. Overall, these tandem and complementary approaches of chemical ecology experiments and plant genomics are crucial in developing insect resistant crops.

PO0697: Fruit Species

Phenotypic Analysis of the U.S. Cucumber PI Core Collection for Fruit Morphological Diversity

Ying-Chen Lin¹, Stephanie Rett-Cadman¹, Yiqun Weng², Zhangjun Fei³ and Rebecca Grumet¹,
(1)Department of Horticulture, Michigan State University, East Lansing, MI, (2)USDA ARS/ University of Wisconsin, Madison, WI, (3)Boyce Thompson Institute, Cornell University, Ithaca, NY

The cucumber (*Cucumis sativus* L.) germplasm plant introduction (PI) accessions from the United States National Plant Germplasm System (NPGS) collection (1,234 accessions) were previously genotyped as part of the CucCAP project using the genotyping-by-sequencing (GBS) technology, and resultant SNP data were used to identify a core collection. This collection of 395 lines represents >95% of the whole allelic diversity, along with accessions possessing key disease resistances, fruit quality and agronomic traits. The core collection is currently being re-sequenced to facilitate single-base resolution genotyping. In the summer of 2019, 263 accessions from the core collection were grown in the field at the MSU Horticulture Teaching and Research Center (HTRC) for phenotypic analyses of fruit characteristics. Fruit were harvested at three stages of development: early exponential growth at 5-7 days post pollination (dpp); post-exponential growth at 16-20 dpp; and maturity at 30-40 dpp. Fruits at 5-7 dpp were photographed and characterized for fruit shape and color, and spine density and color. For fruits at 16-20 and 30-40 dpp, photographs were taken of whole fruit and cross sections sampled from the mid-region of the fruit. The photos were used to analyze length, diameter, shape, skin color, flesh color, flesh thickness, and internal cavity size. The morphological and genotypic data will be used for genome wide association study (GWAS) to provide better understanding of cucumber morphological diversity and provide markers for fruit quality characteristics.

PE0698: Fruit Species

De novo Assembly of *Aronia melanocarpa* Fruit Transcriptome to Identify Genes Involved with Polyphenol Biosynthesis

Jonathan Mahoney¹, Liam Iorio¹, Jill L. Wegrzyn², Bradley W. Bolling³, Mark H. Brand¹ and Huanzhong Wang¹, (1)University of Connecticut, Storrs, CT, (2)Department of Ecology and Evolutionary Biology, University of Connecticut, Storrs, CT, (3)University of Wisconsin, Madison, WI

Native to eastern regions in North America, the genus *Aronia* is a group of deciduous shrubs in the Rosaceae family, subtribe Pyrinae. Interest in *Aronia* fruit has increased because of their high levels of antioxidants and polyphenols and wide adaptability to various geographic regions with few disease and pest issues. Using Illumina RNA-seq, this study investigated the transcriptome of *Aronia* fruit development to understand the molecular mechanisms involved with *Aronia* fruit polyphenol biosynthesis. Six accessions of *A. melanocarpa* (2x) were collected at four developmental stages. Anthocyanin content tended to increase at each developmental stage. *De novo* assembly was performed with 341.75 million clean reads from 24 samples and assembled into 90,008 transcripts with an average length of 801 bp. The transcriptome had 96.1% complete, 2.3% fragmented and 1.6% missing Benchmarking

Universal Single-Copy Orthologs (BUSCOs) and 85 to 88% of the clean reads aligned to the *de novo* transcriptome. The differentially expressed genes (DEGs) identified in a series of six pairwise comparisons between the four developmental stages 0v1, 0v2, 0v3, 1v2, 1v3 and 2v3 were 277, 441, 1572, 18, 169 and 6, respectively. Several of the significant GO Biological terms included flavonoid biosynthetic and metabolic process, pigment biosynthetic process, carbohydrate metabolic process and polysaccharide metabolic process. Of the 5,799 DEGs, four biosynthetic genes were selected as candidate genes in the flavonoid pathway, C4H, DFR, FLS1 and LDOX. All four genes showed significant up-regulation between the four developmental stages. Additionally, several of the DEGs were identified as encoding MYB, bHLH, and WRKY transcription factors. The qRT-PCR expression results were consistent with the RNA-seq results of candidate genes involved with flavonoid biosynthesis. The use of this *de novo* transcriptome data will continue to enhance our understanding of the molecular mechanisms of polyphenol biosynthesis in *Aronia* fruit.

PO0699: Fruit Species

Genome-Wide Study on Genetic Diversity and Phylogeny of 40 Species in Pomegranate

Han Qu, University of California Riverside, Riverside, CA

We report a large-scale analysis of the patterns of genome-wide genetic variation in pomegranate. 40 accessions were sequenced to an average of x16 depth and 87% coverage. Structure and phylogeny analysis were conducted, also we identified 8,056,477 SNPs that may be useful for QTL mapping and association studies. The data here provide a valuable resource for the analysis of pomegranate and to facilitate future breeding and quantitative trait analysis.

PE0700: Fruit Species

Whole Genome Sequencing and Comparative Genomics of *Ceratocystis fimbriata* in Pomegranate

Abhishek Ashok Gowda, University of Horticultural Sciences Bagalkot, India, Bagalkot, India

Pomegranate (*Punicagranatum* L.), an important fruit crop of India. Pomegranate is affected by many serious diseases, among them pomegranate wilt disease caused by *Ceratocystis fimbriata* has been appearing in devastating form. Whole genome sequencing libraries were prepared and sequenced using Illumina NextSeq500 Paired-end sequencing with 150*2. De-novo assembly of Illumina PE data was performed using SPAdes assembler. Gene prediction was carried out using the tool Augustus- 3.1 and a total of 7773 genes were found in the assembled scaffold. Genes were annotated using NCBI BLAST 2.2.29 with the proteins of *Fungi* kingdom taken from Uniprot database. Pathway analysis was done by using KAAS Server using reference fungus organisms. The present genome analysis is used for comparative analysis with that of the draft genome sequence of the available CF (sweet potato isolate) and the other two species of Ceratocystidaceae family, *C. manginecans* and *C. albifundus* and possible identification of isolate/ species specific genes. Comparative pan-genome analysis was performed to study the similarity/diversity within Ceratocystidaceae family members. Results revealed that there were 9659 core genes in all the four isolates and number of unique genes in *C. albifundus* are 2859, *C. fimbriata* (sweet potato)-264, *C. fimbriata* (Pomegranate)-116 and in *C. manginecans*-94. Core genome phylogeny *C. fimbriata* isolate of pomegranate was found to be largely distant from that of the sweet potato isolate. Pan-genome phylogeny confirmed the distinctness of *C. fimbriata* isolate of pomegranate to sweet potato isolate. Interestingly, *C. fimbriata* pomegranate isolate was relatively closer to *C. manginecans* even with pan-genome phylogeny.

PO0701: Fruit Species

A Pangenome for Banana

Habib Rijzaani, Philipp E. Bayer, Soodeh Tirnaz, Jacqueline Batley and David Edwards, University of Western Australia, Perth, WA, Australia

The banana family (Musaceae) is a monophyletic clade of at least two genera, *Musa* and *Ensete*. *Musa*, predominantly grown in Southeast Asia, includes the edible fruit-bearing common banana cultivars, while drought tolerant *Ensete* is predominantly grown in Ethiopia for its starch-rich corm. Here we present a pangenome of banana representing 24 individuals from four *Musa* species and 11 individuals from one *Ensete* species. The resulting pangenome contains 18,359 (39%) core genes and 29,331 (61%) variable genes. Unlike many other plant

pangenomes, the variable genome is not enriched for biotic and abiotic stress resistance, but instead for flower and meristem development, reflecting the morphological differences between the two banana genera. Genetic variability of core genes measured between *Musa* and *Ensete* identified regions associated with flavonoid biosynthesis, defence response, drought tolerance, and metabolite transport. These results support the application of broad diversity for the breeding of more resilient bananas.

PE0702: Fruit Species

In-Vitro Activation of Retro-Transposable Elements As an Effective Mode of Mutagenesis in *Musa Oshry Markovich*, Rahan Meristem (1998) Ltd., Kibutz Rosh-Hanikra, Israel

The majority of triploid banana and plantains are sterile and parthenocarpic. As such they contain a stagnant genome. In recent months we have shown that by demethylation of specific loci in the chromatin it is possible to diversify banana and plantains genomes and modify important characteristics. Using the demethylation compounds 5-Aza-2'-deoxycytidine it is possible to control DNA methylation of plant meristems. This induces activation of retro-transposable elements and consequently generation of an array of new genotypes. Given the new insertions of the retro elements the mutations remain stable for many generations. Employing this technique, we have recently mutated GAL (Cavendish variety) and tested the plants for resistance/tolerance to TR4 (Panama Disease). From a population of 9640 in vitro-mutated plants that have been inoculated with TR4, we have selected 514 lines that exhibited asymptomatic phenotypes. These were recently evaluated in a field trial in an infected area in the Philippines.

An analysis of the mutated genotypes demonstrated sensitivity of particular chromatin regions to the in-vitro demethylation treatment. Using this method of mutagenesis, we have targeted and selected genotypes with the following characteristics: TR4 resistance, Altered plant stature, and Early-flowering.

We have analyzed the polymorphic regions due to demethylation of DNA. We performed an entire genome-sequence analysis between the TR4 resistant genotype and its precursor mother clone. Comparative data show changes in various parts of the genome including in regions that encode resistance genes.

Detailed analysis of the two clones will be discussed in the context of the TR4 resistance.

PO0703: Fruit Species

Genome-Wide Study on Polysomic Genetic Factors Conferring Plasticity of Flower Sexuality in Hexaploid Persimmon

Kanae Masuda¹, Kenta Shirasawa², Eiji Yamamoto², Noriyuki Onoue³, Atsushi Kono³, Koichiro Ushijima¹, Yasutaka Kubo¹, Ryutaro Tao⁴ and Takashi Akagi¹, (1)Okayama University, Okayama, Japan, (2)Kazusa DNA Research Institute, Kisarazu, Japan, (3)Institute of Fruit Tree and Tea Science, National Agriculture and Food Research Organization, Hiroshima, Japan, (4)Graduate School of Agriculture, Kyoto University, Kyoto, Japan

Sexuality is one of the fundamental mechanisms to maintain genetic diversities within a species. Dioecy in diploid persimmons (*Diospyros* spp.) is controlled by the Y chromosome-encoded small-RNA gene, *OGI*, and its autosomal counterpart, *MeGI*. Hexaploid Oriental persimmon (*D. kaki*) has evolved more plastic sexuality based on the same molecular pathways as in diploids, where genetically male individuals, carrying *OGI*, can produce both male and female flowers (monoecy), due to semi-inactivation of *OGI* by the *Kali*-SINE retrotransposon insertion on the promoter region. The mechanism of switching active/inactive states of the *OGI* is, however, little known. Here, to exploit the genetic factors regulating the bias of male and female flowers in hexaploid *D. kaki*, we developed genome-wide association and correlation analytic methods for polyploid genome. Considering the nature of polysomic inheritances, we conducted association/correlation analyses with the following two approaches; (i) treating heterozygosity separately (additive model), and (ii) considering heterozygosity uniformly (binary or dominant model). The quantitative genotypes were determined using allele frequencies in each SNP loci from ddRAD-Seq data, with 83 segregated individuals derived from cv. Yamatogoshu (6A + XXXXXX) x cv. Taishu (6A + YYXXXX). In the Y-chromosomal region including *OGI*, allele dosages were additively correlated with the

biases in flower sexuality. Genome-wide association analysis with binary genotypes, which was compensated with the effect of *OGI* allele dosages, also detected 2 fundamental loci associated with biases in flower sexuality. They included some good candidate genes, especially involving histone remodeling functions, which potentially act for activation of *OGI* expression.

PE0704: Fruit Species

Mutation of a bHLH Transcription Factor Allowed Almond Domestication

Raquel Sánchez-Pérez, CEBAS-CSIC, ESPINARDO (MURCIA), Spain

Authors: Sánchez-Pérez^{1,2,3*}, S. Pavan^{4,5}, R. Mazzeo^{2,3,4}, C. Moldovan^{2,3}, Aiese Cigliano⁶, J. Del Cueto^{1,2,3,7}, F. Ricciardi^{2,3,8}, C. Lotti⁸, L. Ricciardi⁴, Dicenta¹, R. L. López-Marqués⁹, B. Lindberg Møller^{2,3}

¹Department of Plant Breeding, CEBAS-CSIC, Campus Universitario de Espinardo, 30100 Espinardo, Spain.

²Plant Biochemistry Laboratory, Department of Plant and Environmental Sciences, University of Copenhagen, DK-187, Frederiksberg C, Denmark.

³VILLUM Research Center for Plant Plasticity, Department of Plant and Environmental Sciences, University of Copenhagen, DK-1871 Frederiksberg C, Denmark.

⁴Department of Soil, Plant and Food Science, University of Bari “Aldo Moro,” 70126 Bari, Italy.

⁵Institute of Biomedical Technologies, National Research Council (CNR), 70126 Bari, Italy.

⁶Sequentia Biotech SL, Campus de la UAB, 08193 Bellaterra, Barcelona, Spain.

⁷Arboriculture Research Group, Agroscope, Conthey, Switzerland.

⁸Department of the Sciences of Agriculture, Food and Environment, University of Foggia, 71100 Foggia, Italy.

⁹Transport Biology Section, Department of Plant and Environmental Sciences, University of Copenhagen, DK-1871 Frederiksberg C, Denmark.

Wild almond species are toxic and bitter because of the accumulation of amygdalin, a cyanogenic glucoside present in more than 3,000 plant species. The first two steps in the biosynthesis of amygdalin is due to the expression of two P450s (*PdCYP79D16* and *PdCYP71AN24*). Almond domestication was enabled by the selection of sweet almonds lacking of the expression of these two P450s. With the sequencing of the almond genome we have identified a cluster of five *bHLH* genes, where the *Sweet kernel* locus had been previously localized. Functional characterization in *Nicotiana benthamiana* and in *Saccharomyces cerevisiae* revealed that bHLH2 controls the expression of the two P450s. A non-synonymous point mutation (Leu to Phe), in one of the fourteen amino acids implicated in the dimerization of bHLH2, blocks the transcription of *PdCYP79D16* and *PdCYP71AN24*. The immediate consequence is the absence of prunasin, the precursor of amygdalin, in the seed coat or tegument and therefore the lack of amygdalin in the cotyledon, the edible part, which now will be non-bitter and deliciously sweet.

PO0705: Fruit Species

QTL Mapping for Pecan Scab (*Venturia effusa*) Resistance Based on High-Density Genetic Maps in Pecan (*Carya illinoensis*)

Gaurab Bhattarai¹, Shanshan Cao² and Patrick Conner², (1)University of Georgia, Athens, GA, (2)University of Georgia, Tifton, GA

Pecan (*Carya illinoensis* (Wangenh.) K. Koch) is an important commercial fruit crop grown in the United States. Genetic maps are essential tools for the gene positional cloning and marker-assisted breeding. In this study, an F₁ mapping population, derived from a cross of two pecan cultivars ‘Pawnee’ as a female and ‘Elliot’ as a male, were used to construct the linkage based genetic maps of each parent using genotyping-by-sequencing (GBS) method. GBS yielded, on average, 1,817,986 good quality barcoded reads for each sample. The TASSEL5-GBSv2 pipeline was used to call the single nucleotide polymorphism (SNP) using a *C. illinoensis* ‘Elliot’ reference genome sequence (Unpublished). The 8,264 high-quality SNP markers were detected after filtering for several quality parameters. Sixteen linkage groups were then constructed for each parent in JoinMap5.0 following the two-way pseudo-test-cross strategy used for cross-pollinated species. Pawnee and Elliot map contains 1,538 and 1,256 markers spanning 2,335.13 and 2,500 cM with average marker intervals of 1.52 and 1.99 cM, respectively. The high correlation between genetic and physical distances of the markers in both parental maps proves the high synteny between the genetic maps and the reference genome sequence. Percentage of germinated conidia producing subcuticular hyphae within the leaf were counted as a pecan leaf scab disease severity phenotype. We identified a significant QTL on chromosome 5 in Elliot conferring resistance against a ‘Pa-Fh-B’ isolate of the scab pathogen. So far, these are the first high-density linkage maps of pecan constructed and are expected to provide a strong foundation for quantitative trait loci (QTL) mapping and marker-assisted selection for other economically important traits in the near future.

PE0706: Fruit Species

Linkage Mapping and QTL Analysis of Pecan Scab Susceptibility and Spring Emergence from Genotyping By Sequencing (GBS) of a Full-Sibling Pecan Population

Nolan B. Bentley¹, L.J. Grauke², Xinwang Wang², Erin Ruhlman¹, Robert R. Klein³ and Patricia E. Klein⁴, (1)Department of Horticultural Sciences, Texas A&M University, College Station, TX, (2)USDA ARS Pecan Breeding & Genetics, Somerville, TX, (3)USDA-ARS Crop Germplasm Research Unit, College Station, TX, (4)Department of Horticultural Sciences, Texas A&M University, College Station, TX

Pecan (*Carya illinoensis*) is an outcrossing, highly heterozygous, and slow to mature tree native to North America utilized for its culinary, ornamental, and lumber qualities. Recent efforts to expand the genomic resources for this crop have led to the development of chromosome-scale reference genomes (HudsonAlpha, data unpublished). The present research utilized one of these reference genomes and a cost-effective and high-throughput genotyping by sequencing (GBS) technique to discover SNPs segregating in a mapping population of 151 full-sibling progeny of the pecan cultivars ‘Elliott’ and ‘VC1-68’ and map them using a pseudo-testcross mapping strategy. Single-parent informative SNPs (n=5,556) were mapped to 32 linkage groups corresponding to the 16 chromosomes of pecan in each parent with an average of 173.6 SNPs per linkage group and an average spacing of 0.55 cM between SNPs. Total map lengths were 1,484 cM in the ‘Elliott’ map and 1,560 cM in the ‘VC1-68’ map. QTL analysis detected one locus predictive of pecan scab susceptibility across two years of evaluation and two loci predictive of the emergence of leaves in the spring (bud-break) over three years of evaluation. These loci were supported by multiple SNPs with LOD p-values less than 0.05 using classical interval mapping tested against 1,000 permutations of the phenotype. The locus most predictive for bud-break has a large effect (LOD 10.06) and is syntenic with a recently reported region of the English walnut (*Juglans regia*) genome that is similarly predictive. This suggests that this locus may contain a mechanism controlling spring-emergence conserved across the *Juglandaceae*. These loci and genetic maps are valuable as the development of genomic tools capable of predicting plant performance is a priority goal for pecan breeding programs due to pecan’s long juvenile period and variability at major effect loci. Additionally, the genetic maps and techniques developed in this project will directly facilitate the mapping of high-value fruiting traits in the near future as this population completes the transition to sexual maturity.

PO0707: Fruit Species

Loci Determining Trunk Caliper of Interspecific Pistachio Rootstock Identified using Phased, Chromosome-Scale Genome Assemblies

William J. Palmer¹, Ewelina Jaczygrad², Salih Kafkas³, Deborah Golino², John E. Preece⁴ and Richard Micheltore⁵, (1)Genome Center, University of California, Davis, CA, (2)Foundation Plant

Services, Davis, CA, (3)University of Cukurova, Adana, Turkey, (4)Nat'l Clonal Germplasm Repository, Davis, CA, (5)The Genome Center, University of California, Davis, CA

We have developed phased, chromosome-scale genome assemblies for both parents of UCB-1, the most common rootstock used in the \$1.6 billion American pistachio industry. UCB-1 is an interspecific F1 hybrid of *Pistacia atlantica* and *P. integerrima* that was initially bred to confer both Verticillium and frost tolerance. Because these species are divergent, outbreeding, and heterozygous - the resultant F1 seedling rootstocks exhibit considerable variation in both size and form that results in variation in the clonal, grafted scion. In commercial orchards, this size variation often necessitates the costly removal of small individuals and replanting. As the nut-producing scion is typically clonal *P. vera*, the genetic basis of this variation in the rootstock is of great interest for marker-development and selection of uniformly-sized trees prior to planting in orchards.

We have utilized our chromosome-scale assemblies, a mapping population of over 3,000 UCB-1 trees, phenotypic data from multiple years, and a linkage, phasing, and trait association workflow to identify loci strongly associated with trunk caliper in both commercial (grafted) and experimental (ungrafted) orchards. We leveraged short and long-read technologies including 10X Chromium, Dovetail Hi-C, and Oxford Nanopore Promethion as well as GBS and Skim-seq sequencing strategies. The combination of phased, chromosome-scale assemblies of the parental trees, and the very large F1 mapping population from the same two trees allowed us to dissect the basis of trunk caliper and its genotype-by-environment interaction in extremely fine detail. This revealed a major effect locus associated with variation in trunk caliper, together with several others interacting epistatically. Together these loci explain and describe size variation observed in commercial orchards.

PE0708: Fruit Species

Comparative Genomics of Six Juglans Species Reveals Disease-Associated Gene Family Contractions.

Alexander Trouern-Trend, Department of Ecology and Evolutionary Biology, University of Connecticut, Storrs, CT

Juglans, the most speciose genus in the walnut family (Juglandaceae) represents most of the family's commercially valuable fruit and wood-producing trees and includes several species used as rootstock for their resistance to various physical and biological stressors.

We present the full structural and functional genome annotations of six members of *Juglans* and one outgroup within Juglandaceae (*Juglans regia*, *J. cathayensis*, *J. hindsii*, *J. microcarpa*, *J. nigra*, *J. sigillata* and *Pterocarya stenoptera*) using the BRAKER2 semi-supervised gene prediction pipeline and EnTAP. The sizes of these assemblies range between 641 Mb (*J. nigra*) and 992 Mb (*P. stenoptera*). For each of the 7 annotations, gene predictors were trained using 19 tissue-specific *J. regia* transcriptomes, which offered a gradient of genome-evidence phylogenetic distance. Additional evidence (EnTAP pipeline) and filters (gFACs) were applied to multiexonic and monoexonic putative genes predicted by BRAKER2 to yield between 27,000 and 44,000 high-confidence gene models per species. Comparison of gene models to the BUSCO embryophyta dataset suggested that, on average, genome completeness was 85.6%.

We utilized these annotations to assess gene family evolution within *Juglans* and between *Juglans* and selected species from across the breadth of Embryophyta. Finally, an ancient whole genome duplication that took place in a common ancestor of Juglandaceae was dated using site substitution comparative analysis.

PO0709: Fruit Species

Can We Use miRNA to Predict Juvenility and Reproductive Competence in Tree Crops?

Muhammad Umair Ahsan¹, Alice Hayward² and Neena Mitter², (1)Cornell University, Ithaca, NY, (2)The University of Queensland, St Lucia, Australia

Vegetative to reproductive phase transition in plants is regulated by the sequential activity of two microRNAs: miR156 and miR172. A decline in miR156 and increase in miR172 abundance is associated with phase transition.

There is very limited information on phase transition in economically important horticultural tree crops, which have a significantly long vegetative phase affecting fruit bearing. In this study, we profiled various molecular cues known to be involved in phase transition and flowering, including the miRNAs miR156 and miR172, in three horticultural tree crops avocado (*Persea americana*), mango (*Mangifera indica*) and macadamia (*Macadamia integrifolia*). We observed that miR156 expression decreases as these trees age and can potentially be used as a juvenility marker. Consistent with findings in annual plants, we also observed conserved regulation of the miR156-*SPL4* regulatory module in these genetically distant tree crops, suggesting that this pathway may play a highly conserved role in vegetative identity. Meanwhile, the abundance of miR172 and its target *AP2*-like genes, as well as the accumulation level of *SPL9* transcripts, were not related with plant age in these crops except in avocado where miR172 expression increased steadily. Overall, this study provides an insight into the molecular associations of juvenility and phase transition in horticultural trees where crop breeding and improvement is encumbered by long juvenile phases.

PE0710: Fruit Species

The Draft Genome of Dragon Fruit Reveals Gene Families Related to Drought Resistance

Jinfang Zheng¹, Xuehuan Feng¹, Dapeng Zhang² and Yanbin Yin³, (1)University of Nebraska-Lincoln, (2)Sustainable Perennial Crops Laboratory, USDA/ARS, Beltsville, MD, (3)University of Nebraska-Lincoln, Lincoln, NE

Dragon fruit (*Hylocereus undatus*) is a vining cactus native to tropical America. This crop is highly water efficient and adapts very well to southern California, Florida, Hawaii and Puerto Rico. *H. undatus* belongs to the family *Cactaceae*, which has five sequenced genomes in the draft status including *Carnegiea gigantea* (saguaro cactus). Here, we sequenced the genome of *H. undatus*, aiming to understand the genomic basis of drought resistance in *Cactaceae*. By comparing to 12 genomes of C3, C4, and CAM plants, we have identified orthologous protein families that are significantly expanded/enriched in *Cactaceae* and might be related to drought resistance.

The 1,332 Mbp genome of the *H. undatus* cultivar “PuertoRico-1” was sequenced using a combination of 10X Chromium sequencing and long-read sequencing from Pacific Biosciences, Chicago and Hi-C chromatin proximity ligation from Dovetail Genomics. The initial 10X raw read assembly had a contig N50 at 31kb, and scaffold N50 at 769kb. This final assembly was significantly improved from the 10X assembly with Chicago and Hi-C data for scaffolding using the HiRise pipeline, which has scaffold N50 at 110Mb, and N90 at 0.041Mb. It was estimated that over 87% of the dragon fruit genome was assembled into 33,691 scaffolds, with the longest scaffold at 129Mb long, and 90% of the assembled genome was contained in 279 scaffolds and 50% of the assembled genomewas contained in just 6 scaffolds.

BUSCO showed that the dragon fruit draft genome included 93% core eukaryotic core genes. RepeatModeler and RepeatMasker were used to identify repeat regions and masked 58.76% of the genome. An *ab initio* gene prediction using the program MAKER with the help of RNA-seq data predicted 31,520 protein coding genes. tRNA-scan-SE2 and Infernal with Rfam predicted 5,536 tRNA genes, 2,364 rRNA genes, and 5,697 non-coding RNA genes. Orthofinder was used to predict orthologous protein families in *H. undatus* and 12 other plant genomes, which generated 519 single copy orthologous groups (orthogroups). Based on these orthogroups, a species tree was built and the species divergence time was inferred by using r8s. The divergent time of most species were consistent with the timetree (http://www.timetree.org/search/goto_timeline). The divergent time of dragon fruit (*H. undatus*) and cactus (*C. gigantea*) was estimated to be 27.53 MYA. We further conducted an analysis of gene family expansion and contraction. Compared to other plants, the immediate ancestor of dragon fruit and cactus had 58 significantly expanded gene families and 7 significantly contracted gene families. A Gene Ontology enrichment analysis has found that the 58 significantly expanded gene families were enriched with drought resistance-related GO terms, such as stomatal movement, synthesis of protective macromolecules and the synthesis of secondary metabolism. Our research not only reported the draft genome of dragon fruit but also helped understand the drought resistance of these iconic *Cactaceae* plants.

PO0711: Fruit Species

Diversity in Metabolites and Fruit Quality Traits in Blueberry Enables Ploidy and Species Differentiation and Establishes a Strategy for Genetic Studies on Bioactive Traits

Massimo Iorizzo¹, Molla F. Mengist², Mary H. Grace², Jia Xiong², Colin D. Kay², Nahla Bassil³, Kim Hummer³, Mario Ferruzzi² and Mary Ann Lila⁴, (1)Plants for Human Health Institute, Department of Horticultural Science, North Carolina State University, Kannapolis, NC, Madison, WI, (2)Plants for Human Health Institute, North Carolina State University, NCRC, 600 Laureate Way, Kannapolis, NC, 28081, USA, (3)USDA-ARS National Clonal Germplasm Repository, Corvallis, OR, (4)Plants for Human Health Institute, Kannapolis, NC

Blueberry is well recognized as a rich source of health promoting phytochemicals such as flavonoids and phenolic acids. Despite the important roles blueberries have on health effects, information is limited about the levels of variation in bioactive compounds within and between ploidy level and species, and their association with fruit quality traits. Such information is crucial to elucidate the genetic mechanisms controlling biosynthesis of these compounds in blueberry. Hence, our objective was to evaluate 33 phytochemicals belonging to four major groups of flavonoids and phenolic acids across 128 blueberry accessions over two years together with fruit quality traits, including fruit weight, titratable acidity, total soluble acids and pH. Highly significant variation among accessions, years, and accession by year interaction were identified for most of the traits. Broad sense heritability of traits ranged from 20% to 90%, with most traits revealing moderate to high broad sense heritability ($H^2 > 40\%$). Cluster analysis grouped phytochemicals by their functional structure (eg. anthocyanins, flavanols and flavonols). Fruit weight showed a negative correlation with most of the metabolites. Multivariate analysis of the traits resulted in separation of diploid, tetraploid and hexaploid accessions, indicating that each ploidy group has a distinct metabolite profile and a discrete set of fruit quality traits. Overall, traits with high heritability were greatly discriminative, indicating that genotypic effects explain the extensive bioactive and fruit quality trait diversity identified within and between ploidy groups. These results provide a framework to uncover the genetic basis of bioactive and fruit quality traits and will be useful to advance blueberry-breeding programs focusing on these traits.

PE0712: Legumes, Soybean, Common Bean, and related

Legumeinfo.org: Online Resources Facilitate Basic Research and Crop Improvement for Legumes

Wei Huang¹, Joel Berendzen², Anne V. Brown³, Connor T. Cameron², Jacqueline D. Campbell⁴, Ethalinda Cannon³, Alan Cleary², Sudhansu Dash², Samuel A. Hokin², Sven Gato Redsun², Scott R. Kalberer³, Nathan Weeks³, Andrew Wilkey³, Andrew Farmer², Steven B. Cannon³ and LIS group, (1)ORISE Fellow, USDA-ARS-CICGRU, Ames, IA, (2)National Center for Genome Resources (NCGR), Santa Fe, NM, (3)USDA-ARS-CICGRU, Ames, IA, (4)Iowa State University, Ames, IA

The legume family (*Fabaceae*) contains many important forage and food crops. The Legume Information System (LIS; <https://legumeinfo.org>) is a resource for comparative legume research and for assisting researchers in discovering the molecular basis for important traits, and for exploring genomic and genetic data for legume crops and model species. Currently, LIS hosts annotated genomes for 16 legume species. LIS partners with other plant data resources including Araport/JCVI, Cowpea Genome Project, KnowPulse, PeanutBase, Phytozome, Pulse Crop Database, SoyBase and others, providing methods to integrate genomic, genetic and trait data by providing links to other legume data sets not housed at LIS. In the past year, LIS has seen the following improvements, in coordination with LIS partners: a comparative genetic map viewer (CMap-js), a new Genotype Comparison Visualization Tool (GCViT); tools for annotating genes and placing user-submitted genes into the families and trees; InterMine interfaces for seven legume species; numerous improvements to the Genomic Context Viewer (for exploring genomic micro- and macro-synteny), and a new pea (*Pisum sativum*) genome and gene sets. LIS is funded by the USDA-ARS and is jointly developed and maintained by the National Center for Genome Resources (NCGR) and the USDA-ARS at Ames, Iowa. LIS is a member of the AgBioData consortium, which works to create database products that are more FAIR (Findable, Accessible, Interoperable and Reusable); and promotes sharing common resources and data standards.

P00713: Legumes, Soybean, Common Bean, and related

LegumeIP V3: Empowering Comparative Genomics, Transcriptomics, and Phylogenomics in Legumes

Xinbin Dai, Yibo Dong, Zhaohong Zhuang and Patrick X. Zhao, Noble Research Institute, Ardmore, OK

We present the LegumeIP, which is an integrative platform to study gene functions, biological networks, pathways, and genome evolution in legumes. We completely re-developed the new LegumeIP system (LegumeIP V3) with a focus on facilitating and empowering comparative genomics, transcriptomics, and phylogenomics in Legumes.

The LegumeIP V3 includes a total of 783,393 protein-coding genes from 19 species, including 17 legumes and two outgroup reference species. The species in two economically important legume tribes *Dalbergieae* and *Genisteae*, such as peanut and lupin, were added into the LegumeIP database for the first time.

In the gene expression module, we replaced traditional microarray with RNA-seq data that are available in the public domain. Such improvement extends the gene expression function from three model species to all 19 species. The new version accommodates 85 RNA-seq experiments which include 1,694 RNA-seq runs. It is worth mentioning that the gene expression module was designed to be flexible to add new RNA-seq data upon the user's request as long as the data is publicly accessible through the NCBI Sequence Read Archive (SRA) repository.

Furthermore, the gene expression analysis functions were also re-written to enable users to perform integrative and interactive differential expression analysis, metabolic pathway analysis, and co-expression network analysis. Users can now select any subset of RNA-seq data and multiple species of their interest and perform these analyses on-the-fly in the LegumeIP V3. The enhanced analysis functions provide useful and practical tools to discover key genes and regulatory networks in the hosted species and translated into other species through comparative genomics analysis.

The LegumeIP V3 is available at <http://plantgrn.noble.org/legumeIPV3>. This website is free and open to all users, and there is no login requirement.

PE0714: Legumes, Soybean, Common Bean, and related

GCViT: A Genotype Comparison Visualization Tool

Anne V. Brown¹, Andrew Wilkey¹, Nathan Weeks¹, Wei Huang² and Steven B. Cannon¹, (1)USDA-ARS-CICGRU, Ames, IA, (2)ORISE Fellow, USDA-ARS-CICGRU, Ames, IA

As high throughput genotyping costs have dropped, the dense genotyping of large germplasm collections to identify genomic regions associated with particular traits has become feasible. However, due to the vast number of SNPs produced, data analysis can be a challenge. Here we describe GCViT, which is a new tool that allows users to visually compare differences and similarities between two or more accessions using SNP data. GCViT combines a data service along with a graphical user configuration tool. GCViT takes a VCF file and performs a pairwise comparison between the user selected accessions in the dataset that is then plotted on the chromosomes in a short amount of time. Early use shows this tool can be particularly useful for comparing plant breeding lines, and the determination of potentially useful genetic regions with the lines.

PO0715: Legumes, Soybean, Common Bean, and related

Rhizobial tRNA-Derived Small RNAs are Signal Molecules Regulating Plant Nodulation

Bo Ren, Xutong Wang, Jingbo Duan and **Jianxin Ma**, Purdue University, West Lafayette, IN

Rhizobial infection and root nodule formation in legumes requires recognition of signal molecules produced by the bacteria and their hosts. Here we show that rhizobial tRNA-derived small RNA fragments (tRFs) are signal molecules that modulate host nodulation. Three families of rhizobial tRFs were confirmed to regulate host genes associated with nodule initiation and development via hijacking the host RNAi machinery that involves ARGONAUTE 1. Silencing individual tRFs with the use of short tandem target mimics or by overexpressing their targets represses root hair curling and nodule formation, whereas repressing these targets with artificial miRNAs identical to the respective tRFs or mutating these targets with CRISPR-Cas9 promotes nodulation. Our findings thus uncover a bacterial small RNA-mediated mechanism for prokaryote-eukaryote interaction and may pave the way for enhancing nodulation efficiency in legumes.

PE0716: Legumes, Soybean, Common Bean, and related

New Uses and Plant Breeding Targets for Nutritional Benefit in Feeding Animals

Amanda M. Hulse-Kemp¹, Ondulla T Toomer², Lisa L Dean², Deborah L Boykin³, Jeffrey Dunne⁴, Ramon Malheiros⁴ and Kenneth Anderson⁴, (1)USDA-ARS-GBRU, Raleigh, NC, (2)USDA-ARS, Raleigh, NC, (3)USDA-ARS, Stoneville, MS, (4)North Carolina State University, Raleigh, NC

Previous studies have identified normal-oleic peanuts as a suitable and economical broiler feed ingredient. However, no studies to date have examined the use of high-oleic (HO) peanut cultivars as a feed ingredient for laying hens and determined the impact of feeding HO peanuts on performance and egg nutritive qualities. We aimed to examine the use of HO peanuts as a feed ingredient for layer hens to determine the effect on performance, egg lipid chemistry, and quality of the eggs produced. No significant differences in hen performance or egg quality as measured by USDA grade quality, egg albumen height, or egg Haugh unit between the treatment groups. However there were significant differences in many beneficial traits of interest fed the peanut diet, including lower egg weights, greater palmitic and stearic acid, lower trans-fat and yellower egg yolks. All egg protein extracts from all treatments at each time point were non-reactive with rabbit anti-peanut agglutinin antibodies. The results of the study indicate that HO peanuts could be used to produce beneficial traits in poultry. As an important commodity crop in the southeastern United States, production areas closely overlap those of the poultry industry and could be of economic advantage to producers while providing a potential health benefit to the consumer with improved egg nutrition. As results are promising and of great interest to the poultry industry, this is a unique opportunity to target nutritional profiles of the commodity via plant breeding in a collaborative manner to optimize the HO, protein, and fat components in the peanut for additional use in feeding animals.

PO0717: Legumes, Soybean, Common Bean, and related

SMRT- RenSeq for NLR Resistance Gene Characterization in *Arachis*, *Glycine* and *Musa* Species

Gabriel S. C. Alves¹, Ana Zotta Mota², Fernando C. A. Fonseca¹, Amanda Cristina Araújo¹, Michelle Guittton Cotta¹, Roberto C Togawa², Marcos M. C. Costa², Orzenil B. da Silva Junior², Fabricio B.M. Arraes², Maria F. Grossi de Sa², Ana C. M. Brasileiro², Jacob Enk³, **Tianying Lan**³, Patricia M. Guimarães² and Robert N. G. Miller¹, (1)University of Brasília, Brasília, Brazil, (2)Embrapa Genetic Resources & Biotechnology, Brazil, (3)Arbor Biosciences, Ann Arbor, MI

The most abundant plant disease resistance (*R*) genes comprise a single family that encode intracellular multidomain receptor proteins with a stereotypical nucleotide binding site (NBS) and leucine-rich repeat (LRR) domain (NLRs). Conserved motifs in such receptors across diverse plant taxa offer a means for their accelerated isolation across target or uncharacterized plant species. Conserved motifs in domains in NLR genes from reference genomes *Musa acuminata* ssp. *malaccensis* (DH-Pahang V2), *Arachis duranensis* (Aradu 1.0) and *Glycine max* (Williams 82) were used to build species-specific HMM profiles for screening of genomic and transcriptomic sequence data for pfam NLR signature domains (TIR, CC, RPW8, NB-ARC and LRR) from the target disease resistant genotypes *Musa acuminata* ssp. *burmannicoides* var. *Calcutta 4*, *Glycine max* (PI595099) and in the wild species *Arachis stenosperma*. Using specific biotinylated RNA baits designed for target enrichment from gDNA, a combined approach of R-gene sequence enrichment and single-molecule real-time sequencing (SMRT RenSeq) is ongoing for accurate sequencing and *de novo* assembly of NLR gene repertoires in the resistant materials. Full length NLR genes are applicable for functional gene cloning for engineering of durable disease resistance.

PE0718: Legumes, Soybean, Common Bean, and related

A Map of Genetic Variation from 781 Soybean Genomes

Myung-Shin Kim¹, Roberto Lozano², Ji Hong Kim³, Dong Nyuk Bae³, Man Soo Choi⁴, Namhee Jeong⁴, Soo-Kwon Park⁵, Michael A. Gore⁶, Jung-Kyung Moon⁷ and **Soon-Chun Jeong**³, (1)Department of Plant Science, Plant Genomics and Breeding Institute and Interdisciplinary Program in Agricultural Genomics, Seoul National University, Seoul, Korea, Republic of (South), (2)Cornell University, Ithaca, NY, (3)Korea Research Institute of Bioscience and Biotechnology, Chungbuk, South Korea, (4)National Institute of Crop Science, Jeonbuk, South Korea, (5)National Institute of Crop Science, Rural

Development Administration, Jeonju, Korea, Republic of (South), (6)Plant Breeding and Genetics Section, School of Integrative Plant Science, Cornell University, Ithaca, NY, (7)National Academy of Agricultural Sciences, Jeonbuk, South Korea

Soybean is an economically and environmentally important major crop worldwide. It is a predominant plant protein and oil source of both food and feed and has capacity to fix atmospheric nitrogen by intimate symbioses with microorganisms. Here we present a fine genome-wide variation map in 781 accessions including 418 domesticated (*Glycine max*) and 345 wild (*Glycine soja*) soybeans and 18 of their natural hybrids. We identified 31 million single nucleotide polymorphisms and 5.7 million small indels that contribute to within- and between-population variation. We describe a comprehensive characterization of the geographic and functional differentiation of rare and common genetic variants with insights into the domestication history of soybean and detection of domestication-selective sweeps. We show that this resource enables us to increase marker density of existing data sets for improving the resolution of association studies.

PO0719: Legumes, Soybean, Common Bean, and related

Chromosomal Features and Rearrangements Revealed by Comparison of Genetic Maps of *Glycine max* and *Glycine soja*

Soon-Chun Jeong¹, Kwanghee Lee¹ and Myung-Shin Kim², (1)Korea Research Institute of Bioscience and Biotechnology, Chungbuk, South Korea, (2)Department of Plant Science, Plant Genomics and Breeding Institute and Interdisciplinary Program in Agricultural Genomics, Seoul National University, Seoul, Korea, Republic of (South)

Recombination is a crucial component of evolution and breeding. New combinations of variation on chromosomes are shaped by recombination. Recombination is also involved in chromosomal rearrangements. However, recombination rates vary tremendously among chromosome segments. Genome-wide genetic maps are one of the best tools to study variation of recombination. Here, we describe high density genetic maps of *Glycine max* and *Glycine soja* constructed from four segregating populations. The maps were used to identify chromosomal rearrangements and find the highly predictable pattern of cross-overs on the broad scale in soybean. Markers on these genetic maps were used to evaluate assembly quality of the current soybean reference genome sequence. We find a strong inversion candidate larger than 3 Mb based on patterns of cross-overs. We also identify quantitative trait loci (QTL) that control number of cross-overs. This study provides fundamental insights relevant to practical strategy for breeding programs and for pan-genome researches.

PE0720: Legumes, Soybean, Common Bean, and related

Soybean Haplotype Map (GmHapMap): A Universal Resource for Soybean Translational and Functional Genomics

Francois Belzile, Laval University, Quebec, QC, Canada

The advent of next-generation sequencing (NGS) technologies has provided an exceptional opportunity to systematically detect both nucleotide and haplotype diversity in plant and animal genomes. Here we describe a worldwide haplotype map for soybean (GmHapMap) constructed using whole-genome sequence data for 1,007 *Glycine max* accessions and yielding 15 million variants. The number of unique haplotypes plateaued within this collection (4.3 million tag SNPs) suggesting extensive coverage of diversity within the cultivated germplasm. GmHapMap variants were imputed onto 21,618 previously genotyped (50K array/210K GBS) accessions with up to 96% success for common alleles. A GWAS performed with imputed data enabled us to identify a candidate causal SNP residing in the *NPC1* gene and to demonstrate its role in controlling seed oil content. We determined gene-centric haplotypes (405,101 GCHs) for the 55,589 genes and show that such haplotypes can help to define alleles and 353 genes with a unique haplotype were spotted in highly constrained genic regions. Finally, we predicted 18,031 putative loss-of-function (LOF) mutations in 10,662 genes and illustrate how such a resource can be used to explore gene function. The GmHapMap provides a unique worldwide resource for applied soybean genomics and breeding.

PO0721: Legumes, Soybean, Common Bean, and related

Estimating the Sensitivity and Precision of Structural Variant Discovery From Short Reads Using Long-Read Sequencing in Soybean

Marc-André Lemay¹, Jonas Andreas Sibbesen^{2,3}, Davoud Torkamaneh⁴, Jérémie Hamel⁴, Roger C. Lévesque⁴, Anders Krogh² and Francois Belzile¹, (1)Laval University, Quebec, QC, Canada, (2)Bioinformatics Centre, University of Copenhagen, Copenhagen, Denmark, (3)UC Santa Cruz Genomics Institute, Santa Cruz, CA, (4)Université Laval, Québec, QC, Canada

Despite the increased recognition of structural variants (SVs) as a key component of genetic variation in crops, methods for their discovery and genotyping from high-throughput sequencing data have lagged behind those for single-nucleotide variants (SNVs). Benchmarks of SV discovery performance from short reads consistently document low sensitivity and precision, yet such studies are scarce for plant genomes, which are challenging due to repeat content and past whole-genome duplication events. Long-read sequencing provides a more reliable way of calling SVs, but these methods remain cost-prohibitive on a large scale. Short read-based approaches are therefore likely to remain useful for population-scale SV calling in the near future, and as such, optimizing their analysis in crop genomes is critical. In this study, we generated a dataset of structural variants found in 33 Canadian soybean genomes using 5 to 35X coverage Illumina paired-end reads. Structural variants were discovered using a combination of mapping- and assembly-based approaches and genotyped in five samples using BayesTyper. In parallel, we produced 10 to 15X coverage Oxford Nanopore long-read sequencing data for these five samples in order to generate a reference structural variant dataset against which to benchmark the short read-based methods. Structural variants were called from Oxford Nanopore data by aligning the reads to the reference genome using NGMLR and calling structural variants using Sniffles. Sensitivity and precision of short-read approaches were estimated by comparison to the long-read dataset and allowed the identification of filtering parameters to generate a high-confidence SV dataset from Illumina sequencing data.

PE0722: Legumes, Soybean, Common Bean, and related

Developing a Cost-Effective Strategy for Implementing Genomic Selection in a Soybean Breeding Program

Francisco Ernesto Gomez, University of Minnesota, Saint Paul, MN and Aaron Lorenz, Department of Agronomy and Plant Genetics, University of Minnesota, St. Paul, MN

Genomic selection (GS) has great potential to increase the scale, speed and accuracy of breeding operations and promises an increase rate of genetic gain while also re-shaping breeding programs. A recently proposed two-part GS program aims at increasing genetic gains while being cost-effective. The two-part program reorganizes a conventional breeding program into two distinct parts: population improvement through rapid recurrent GS to increase mean value of the population, and a product development component comprised of a conventional breeding scheme to identify and release new varieties. While the proposed two-part GS strategy was based on the structure of a wheat breeding program, here we evaluate the two-part GS strategy in a soybean breeding program with different considerations in program structure and logistics (e.g. seed numbers, ease of crosses, lack of DH). Evaluating a two-part strategy in soybean breeding would be informative for the integration of GS into a soybean breeding program. Our objectives were to: (i) use simulations to assess and evaluate expected genetic gain for alternative breeding strategies enabled by GS including a two-part breeding strategy in a soybean breeding program; and (ii) compare the cost effectiveness of all breeding strategies with equal annual operating costs and program's overall performance. Preliminary results demonstrate that using GS in a soybean breeding program increases genetic gain over conventional methods. This study will allow to develop a successful breeding strategy for prediction of soybean yield that is based on resource allocation and will open many opportunities for enhancing productivity in soybean breeding programs.

PO0723: Legumes, Soybean, Common Bean, and related

Genomic Diversity of the Naro Soybean Mini-Core Collection

Hideki Nagasaki¹, Hiromi Kajiya-Kanegae², Akito Kaga³, Ko Hirano⁴, Eri Ogiso-Tanaka⁵, Makoto Matsuoka⁶, Masao Ishimoto⁷, Masatsugu Hashiguchi⁸, Hidenori Tanaka⁸, Ryo Akashi⁸, Hiroyoshi Iwata⁹ and Sachiko Isobe¹⁰, (1)Kazusa DNA Research Institute, Kisarazu, Japan, (2)Research Center for

Agricultural Information Technology, NARO, Tsukuba, Japan, (3)National Institute of Agrobiological Sciences, Tsukuba, Japan, (4)Bioscience and Biotechnology Center, Nagoya University, Nagoya, Japan, (5)National Agriculture and Food Research Organization, Ibaraki, Japan, (6)Bioscience and Biotechnology Center, Nagoya University, Nagoya, Aichi, Japan, (7)National Agriculture and Food Research Organization, Tsukuba, Japan, (8)University of Miyazaki, Miyazaki, Japan, (9)Department of Agricultural and Environmental Biology, The University of Tokyo, Tokyo, Japan, (10)Kazusa DNA Research Institute, Kisarazu, Chiba, Japan

Soybean, *Glycine max* (L.) Merr., is the most important legume globally and has been used as a major source of nutritious feed for humans and livestock. Soybean originated in Southeast Asia and has been important traditional foods in many Asian countries. Soybean mini-core collections that cover diversity of the world and Japanese germplasm has been developed by National Agriculture and Food Research Organization (NARO) (Kaga *et al.* 2012). To survey genetic diversity of the soybean mini-core collections, we performed whole genome resequencing on total 198 accessions; 192 accessions from the mini-core collections, three Japanese cultivars (Misuzudaizu, Norin2, Houjaku Kuwazu), one American cultivar (5002T), one accession of *Glycine soja* (B01167; Collected from Northern Japan), and one cultivar of India (C1329) by Illumina HiSeq with pair-end method. The mapping and variant calls was performed by BWA (Li and Durbin 2009) and GATK (McKenna *et al.*, 2010) on the Gmax_275_v2.0 soybean reference sequence from Phytozome (<https://phytozome.jgi.doe.gov>). Copy Number Variations (CNV) was determined by using CNV-Seq (Xie *et al.* 2009). In order to characterize structural variations (SVs), PacBio whole-genome sequencing was performed for ten representative soybean accessions from these accessions by PacBio Sequel. The long reads were mapped to the reference by using NGMLR and SV were detected by SAMtools and Sniffles (Sedlazeck *et al.*, 2018).

We obtained 10,116,707 SNPs and 2,835,680 indels from SNP annotations by SnpEff (Cingolani *et al.* 2012). The cluster analysis revealed 198 accessions converged into three main phylogenetic groups, which were Primitive group included *G. soja* B01167, Peking (GmWMC084), and 18 other accessions; World group included Williams 82 (GmWMC115), 5002T, and other 56 accessions; Japan groups included 120 accessions from Japan and Korea. The accessions from the same geographical region tended to cluster together. There were two variation on genes with a significant association to flowering date by Genome-wide Association Study (GWAS). Furthermore, the total length of subreads from PacBio were ranged from 10.3~17.7 Gb representing $9.2 \times \sim 15.9 \times$ of the soybean genome. The numbers of SVs and pi values tended to be higher in regions where gene density were high, particularly on chr01, chr05 and chr12. This tendency was clearer in the world and Japan groups. The number of SVs and pi values showed positive correlation in most of regions. Especially, there are some insertions/deletions in *I* locus, which are the cluster region of chalcone synthase (CHS) genes that influence their seed coat pigmentation of soybeans. In contrast, CNV between three groups suggested that only a few large scale genome rearrangements have been occurred during the domestication. In contrast, a large number of sequence variations were identified in the 198 accessions. It was therefore considered that diversification during the domestications through the propagations was mostly occurred by variations such as SNPs and Indels.

PE0724: Legumes, Soybean, Common Bean, and related

High-Resolution Mapping of Recombination Hotspots in Soybean [*Glycine max* (L.) Merr.]

Samantha McConaughy¹, Keenan Amundsen¹, Qijian Song², Vincent Pantalone³ and David L. Hyten¹, (1)University of Nebraska-Lincoln, Lincoln, NE, (2)USDA-ARS, Soybean Genomics and Improvement Lab, Beltsville, MD, (3)University of Tennessee, Knoxville, TN

The resolution to detect recombination hotspots is limited by population size and marker distribution however, in soybean [*Glycine max* (L.) Merr.] a high-density genotyping array, SoySNP50K, was employed on two biparental populations: Williams 82 x Essex consisting of 922 recombinant inbred lines (RILs) with 11,922 polymorphic SNPs and Williams 82 x PI479752 (*G. soja*) containing 1038 RILs and 21,478 polymorphic SNPs. We identified 412 recombination hotspots including 19 common hotspots between the two biparental populations. The majority of the hotspots reside in euchromatin regions however, 9 of the hotspots are located in heterochromatic regions, despite recombination being repressed in the majority of heterochromatin. Genomic associations with hotspots are similar to human, dog, wheat, *Drosophila*, and *Arabidopsis* with a CCN repeat motif, a poly-A stretch motif, and

transcriptional factory activity (~83%). Notably, however, we found enrichment with the *Tourist* family of mini inverted-repeat transposable elements (MITEs) that resides in 0.33% of the soybean genome.

PO0725: Legumes, Soybean, Common Bean, and related

Generating High Density, Low Cost Genotype Data in Soybean [*Glycine max* (L.) Merr.]

Mary Happ, University of Nebraska-Lincoln, Lincoln, NE

Obtaining genome-wide genotype information for millions of SNPs in soybean [*Glycine max* (L.) Merr.] often involves completely resequencing a line at 5X or greater coverage. Currently, hundreds of soybean lines have been resequenced at high depth levels with their data deposited in the NCBI short read archive. This publicly available dataset may be leveraged as an imputation reference panel in combination with skim (low coverage) sequencing of new soybean genotypes to economically obtain high-density SNP information. Ninety-nine soybean lines resequenced at an average of 17.1X were used to generate a reference panel, with over 10 million SNPs called using GATK's Haplotype Caller tool. Whole genome resequencing at approximately 1X depth was performed on 114 previously ungenotyped experimental soybean lines. Coverages down to 0.1X were analyzed by randomly subsetting raw reads from the original 1X sequence data. SNPs discovered in the reference panel were genotyped in the experimental lines after aligning to the soybean reference genome, and missing markers imputed using Beagle 4.1. Sequencing depth of the experimental lines could be reduced to 0.3X while still retaining an accuracy of 97.8%. Accuracy was inversely related to minor allele frequency, and highly correlated with marker linkage disequilibrium. The high accuracy of skim sequencing combined with imputation provides a low cost method for obtaining dense genotypic information that can be used for various genomics applications in soybean.

PE0726: Legumes, Soybean, Common Bean, and related

A GWAS Meta-Analysis Approach to Deconstruct Sources of Genotype-Phenotype Associations in Soybeans

Paul I. Otyama, Iowa State University, Ames, IA, Anne V. Brown, USDA-ARS, Ames, IA, Teshale Mamo, ORISE Fellow, USDA- Agricultural Research Service, Corn Insects and Crop Genetics Research Unit, Ames, IA, 50011, USA., Ames, IA and Steven Cannon, USDA–Agricultural Research Service, Corn Insects and Crop Genetics Research Unit, Ames, IA

A large number of genetic variants are known to control many of the plant traits of ecological and agricultural importance. These variants are spread throughout the genome – each contributing in small ways to the overall expression of the trait. Genome-wide association studies have become a great tool, and thus far, have revealed a great deal about the genetic architecture of these traits.

Even then, we do not know the complete genetic architecture of most of these traits despite several GWAS attempts. More genetic variants delineating a trait's genetic variation, continue to be discovered. This is often a result of large studies with hundreds of accessions genotyped with thousands, if not millions, of genetic variants. Controlling for confounding remains a critical step in obtaining high quality associations, ready to use in crop improvement endeavors. And to that end, several methods have been developed able to control for false associations due to genetic structure and kinship among study individuals.

We show how a meta-analysis approach, utilizing models controlling for structure and kinship, out-performs other methods in revealing the full genetic architecture underlying the control of days to anthesis and days to maturity in soybeans planted across different environments.

PO0727: Legumes, Soybean, Common Bean, and related

Genomic Introgression through Interspecific Hybridization Counteracts Genetic Bottleneck during Soybean Domestication

Xutong Wang, Purdue University, West Lafayette, IN

Evidence of introgression, the transfer of genetic material, between crops and their wild relatives through spontaneous hybridization and subsequent backcrossing has been documented; however, the evolutionary patterns and consequences of introgression and its influence on the processes of crop domestication and varietal diversification are poorly understood. We investigate the genomic landscape and evolution of putative crop-wild-relative introgression by analyzing the nuclear and chloroplast genomes from a panel of wild (*Glycine soja*) and domesticated (*Glycine max*) soybeans. Our data suggest that naturally occurring introgression between wild and domesticated soybeans was widespread and that introgressed variation in both wild and domesticated soybeans was selected against throughout the genomes and preferentially removed from the genomic regions underlying selective sweeps and domestication quantitative trait locus (QTL). In both taxa, putative introgression was preferentially retained in recombination-repressed pericentromeric regions that exhibit lower gene densities, reflecting potential roles of recombination in purging introgression. Despite extensive removal of introgressed variation by recurrent selection for domestication-related QTL and associated genomic regions, spontaneous interspecific hybridization during soybean domestication appear to have contributed to a rapid varietal diversification with high levels of genetic diversity and asymmetric evolution between the nuclear and chloroplast genomes.

PE0728: Legumes, Soybean, Common Bean, and related

Lipidomic and Transcriptomic Profiling of Developing Nodules Reveal the Essential Roles of Active Glycolysis and Fatty Acid and Membrane Lipid Biosynthesis in Soybean Nodulation

Jian Zhao, Anhui Agricultural University, Hefei, China, Gaoyang Zhang, Huazhong Agricultural University, Wuhan, China and Beibei Chen, Huazhong Agricultural University, Wuhan, China

Gaoyang Zhang¹, Beibei Chen², Muhammad Z. Ahmad¹, Sehrish Manan², **Jian Zhao**¹

1 State Key Laboratory of Tea Plant Biology and Utilization, College of Tea and Food Science and Technology, Anhui Agricultural University, Hefei, 230036,

2 National Key Laboratory of Crop Genetic Improvement, Huazhong Agricultural University, Wuhan, 430070 China

Abstract:

Symbiotic rhizobia-legume interactions are energy-demanding processes, and the carbon supply from host cells that is critically required for nodulation and nitrogen fixation is not fully understood. The investigated lipidomic and carbohydrate profiles of the transcriptome of developing nodules revealed highly activated glycolysis, fatty acid (FA), 2-monoacylglycerol (2-MAG), and membrane lipid biosynthesis and transport during nodule development. RNA-Seq profiling of metabolic genes in roots and developing nodules highlighted particularly upregulated pathways involved in biosynthesis and transport of FAs, membrane lipids, and 2-MAG in rhizobia-soybean symbioses via the GRAS-WRI-FatM-GPAT-STR pathway, similar to that in legume-arbuscular mycorrhizal fungi symbiosis. The essential roles of the metabolic pathway during soybean nodulation were further supported by analysis of transgenic hairy roots overexpressing soybean GmWRI1b-OE or-KD and GmLEC2a-OE. GmLEC2a-OE hairy roots produced fewer nodules, in contrast to GmWRI1b-OE hairy roots. GmLEC2a-OE hairy roots displayed different or even opposite expression patterns of the genes involved in glycolysis and the synthesis of fatty acids, 2-MAG, TAG, and membrane lipids compared to GmWRI1b-OE hairy roots. Glycolysis, fatty acid and membrane lipid biosynthesis were repressed in GmLEC2a-OE but increased in GmWRI1b-OE hairy roots, which may account for the reduced nodulation in GmLEC2a-OE hairy roots but increased nodulation in GmWRI1b-OE hairy roots. GmWRI1b-KD hairy roots with reduced FA, membrane lipids (PC and MGDG), and TAG/MAG, also had reduced nodule numbers. These data show that active fatty acid, MAG and membrane lipid biosynthesis are essentially required but that TAG biosynthesis is not essential for nodulation and rhizobia-soybean symbioses. These data shed more light on root lipid metabolism as a prerequisite for soybean nodulation, laying foundations for future detailed investigations of soybean nodulation.

References:

Chen, B., Zhang, G., Li, P., Yang, J., Guo, L., Benning, C., Wang, X., and Zhao, J. (2019) Multiple GmWRI1s are redundantly involved in seed filling and nodulation by regulating plastidic glycolysis, lipid biosynthesis and hormone signalling in soybean (*Glycine max*). *Plant Biotechnol J.* doi: 10.1111/pbi.13183.

Ahmad, M.Z., Rehman, N., Yu, S., Zhou, Y., Haq, B., Li, P., Wang, J., Zeng, Z., and Zhao, J. (2019) GmMAX2-D14 and -KAI interactions-mediated SLs and KAR signaling pathway play essential roles in soybean nodulation and nodule evolution in legume. *Plant J.* doi.org/10.1111/tpj.14545

Zhang G, Bahn SC, Wang G, Zhang Y, Chen B, Zhang Y, Wang X, **Zhao J. (2019)** PLD α 1-knockdown soybean seeds display higher unsaturated glycerolipid contents and seed vigor in high temperature and humidity environments. *Biotechnology for Biofuels*. 12:9. doi: 10.1186/s13068-018-1340-4.

Manan, S., Ahmad, M.Z., Zhang, G., Chen, B., Haq, B.U., Yang, J. and Zhao, J. (2017a) LEAFY COTYLEDON 2 regulates subsets of genes involved in controlling the biosynthesis and catabolism of seed storage substances and seed development. *Front. Plant Sci.* **8**, 1604.

PO0729: Legumes, Soybean, Common Bean, and related

QTL-Fine Mapping of Protein Content in Soybean Grain

Luciano C. Silva¹, Arthur M. A. Bernardeli², Rodrigo M. Lorenzoni², Rafael S. Aguiar², Rafael D. Bueno², Newton D. Piovesan² and Maximillier Dal-Bianco², (1)SAS Institute, Sete Lagoas-MG, Brazil, (2)Universidade Federal de Viçosa, Viçosa-MG, Brazil

In a previous study, a genomic region of approximately 780kb harboring QTL for protein content in soybean grain was identified with SNPs. In this study, we investigate this same genomic region by adding microsatellite markers. Two hundred and fifty-one recombinant inbred lines were evaluated in field trials with replicates in two distinct regions of Brazil in two years, and grain protein was determined by near-infrared spectroscopy. Then, we mapped QTL in each location-year separately, jointly, and using the average traits using the methods of multiple interval mapping (MIM), composite interval mapping (CIM), and multiple regression (MR). While in the MIM and CIM both marker genotypes and pseudo-marker genotypes were used as regressors, in the MR only marker genotypes were used as regressors. Results of MIM, CIM, and MR agreed in pinpointing two genomic regions with statistical significance either by the F-test or the likelihood ratio test. One genomic region with approximately 143kb with strong statistical significance was present in a single location-year combination, while the other genomic region with approximately 190kb had a strong signal in all analyses (separate, joint, and average). To further elucidate the presence of putative genes in these two genomic regions, RNA-seq measurements have been taken on the two parental lines and analyzes are ongoing. Furthermore, near inbred lines for a SNP in the genomic region of 143kb are being evaluated to test the hypothesis that there is a genetic effect in this genomic region.

PE0730: Legumes, Soybean, Common Bean, and related

Seed Protein, Oil, and Isoflavone Contents in ‘Williams 82’ By ‘Forrest’ Recombinant Inbred Line Population of Soybean

Jiazheng (John) Yuan¹, Naoufal Lakhssassi², Nacer Bellaloui³, Frances Betts⁴, Teresa Register⁴, Earl Williams⁴, Tri D. Vuong⁵, Henry T. Nguyen⁵, Khalid Meksem² and **Abdelmajid Kassem**¹, (1)Department of Biological and Forensic Sciences, Fayetteville State University, Fayetteville, NC, (2)Southern Illinois University at Carbondale, Carbondale, IL, (3)USDA-ARS, Crop Genetics Research Unit, Stoneville, MS, (4)Fayetteville State University, Fayetteville, NC, (5)University of Missouri, Columbia, MO

In soybean [*Glycine max* (L.) Merr.], seed protein, oil, and isoflavone contents fluctuate from year to year and depend on the genotype and environmental conditions. Soybean cultivar Forrest was the first cultivar to be identified to possess soybean cyst nematode (SCN) resistance associated with high yield. It was also used to study the sudden death syndrome (SDS) of soybean. Soybean cultivar Williams 82 contains *Phytophthora sojae* resistance gene (*Rps1k*). The cultivar was intensively investigated and the whole genome sequence of Williams 82 was completed

and published in 2010. The objectives of this study were to quantify seed protein, oil, and isoflavone (daidzein, glycitein, and genistein) in 'Forrest' by "Williams 82" recombinant inbred line (RIL) population grown in a field in Spring Lake, NC; to compare these traits in RILs and parental lines; and to study correlations among these traits. The results showed that the RIL population had a mean protein content of 44.5% and a mean oil content of 21.56%. Means for daidzein, glycitein, and genistein contents were 303.83 µg/g, 493.6 µg/g, and 292.06 µg/g, respectively. All the traits studied here were normally distributed in the RIL population. The RILs' mean protein content is higher than the means of both parents (Williams 82 and Forrest). Williams 82's mean oil content is higher than both the means of the RILs and Forrest oil contents. Likewise, Williams 82's means of daidzein, glycitein, and genistein contents are higher than the mean contents of both the RILs and Forrest. Forrest had the lowest protein, oil, and isoflavone contents. Protein content is negatively correlated with all other traits compared. Oil content is positively correlated with isoflavone contents and negatively correlated with protein content. Daidzein content is positively correlated with both glycitein and genistein contents. Glycitein content is positively correlated with genistein content. The results are useful to identify soybean lines with high protein, oil, and isoflavone contents as well as to genetically map QTL for these traits in the future.

PO0731: Legumes, Soybean, Common Bean, and related

Identification of Candidate Genes Controlling Soybean Seed Oil and Protein Content

Xu Mengge, Nanjing Agricultural University, Nanjing, China

Soybean is an important crop providing plant oil and protein, therefore, improving seed oil and protein content has been a major target in soybean breeding program. In this study, our objective is to fine map the QTL controlling soybean seed oil and protein content and identify candidate genes underlying these QTL. A population composed of 300 recombinant inbred lines (RILs) derived from PI595843 with high oil content (HOC) but low protein content (LPC) and WH with low oil content (LOC) but high protein content (HPC) was grown in four environments, and the seed oil and protein content was measured by a Near Infrared (NIR) Instrument. A genetic map was constructed using 4702 bin markers by re-sequencing technology. A total of 27 and 29 QTLs were detected for seed oil and protein content, respectively. The major QTL, *qOil-20-3* and *qProt-20-5*, which explained 26.46% and 28.95% of the phenotypic variation, respectively, were mapped into a same region on Chromosome 20, which could be a pleiotropic locus. A total of 130 candidate genes were annotated in this region, and 93 of them were expressed in seeds, with 23 genes were highly expressed in seeds. These 23 genes were cloned from the two parental lines and sequenced, and four genes showed polymorphism between two parents in the coding sequence (CDS) and resulted in amino acid changes. Real-time RT-PCR analysis further help us identify the candidate genes which are likely involved in controlling soybean seed oil and protein content. Our results will serve as a basis for molecular breeding of soybean cultivars with high seed oil and protein content.

PE0732: Legumes, Soybean, Common Bean, and related

Tilling By Exome Capture Sequencing (TbyECS) to Identify Alleles Involved in Soybean Protein Content

Abdelhalim El Baz, Department of Plant Soil and Agricultural Systems, SIUC, Carbondale, IL and Mallory A Cullen, Naoufal Lakhssassi, Zhou Zhou and Khalid Meksem

TILLING (Targeting Induced Local Lesions in Genomes) is a powerful method that proved its efficiency to study genes functions in many species, the method is based on the screening of large mutagenized populations for mutations in targeted genes. In our study an Ethyl methanesulfonate (EMS)-mutagenized population was screened using TILLING by Exome Capture Sequencing (TbyECS) to identify mutants in 20 genes involved in protein biosynthesis and protein content of soybean such as Glycinin and β-conglycinin.

Bioinformatic tools are necessary to analyze the resulting data and speed the discovery of beneficial alleles by reducing the time needed to identify mutants in our population, web free tools are available to assist in SNP discovery, however, performing the data analysis step by step manually is still time consuming. Therefore, we developed a program using R language, to reduce the analysis of the VCF file and mutants identification time from hours to minutes.

PO0733: Legumes, Soybean, Common Bean, and related

Correlation Analysis of the Essential Amino Acids with Other Quality Traits in Soybean Mutant FM6-847

Jiazheng (John) Yuan¹, Naoufal Lakhssassi², Nacer Bellaloui³, Safiah Kassem⁴, Zakariah Kassem⁴, Sonia Viera¹, Shontia Blount¹, Layla Rashad¹, Daniel Okunbor⁵, Khalid Meksem² and My Abdelmajid Kassem¹, (1)Department of Biological and Forensic Sciences, Fayetteville State University, Fayetteville, NC, (2)Southern Illinois University at Carbondale, Carbondale, IL, (3)USDA-ARS, Crop Genetics Research Unit, Stoneville, MS, (4)Athens Drive High School, Raleigh, NC, (5)Department of Mathematics and Computer Science, Fayetteville State University, Fayetteville, NC

The nine amino acids (phenylalanine, valine, threonine, tryptophan, methionine, leucine, isoleucine, lysine, and histidine) can't be synthesized de novo by human or animal cells and therefore, these amino acids must be obtained from diet. The objective of this study was to analyze the intra-correlation among essential amino acid contents in soybean seeds and decipher the relationship between essential amino acid contents with other quality traits such as protein, oil, and sugar contents in soybean mutant FM6-847 and three USDA checks LD00-2817, LD06-7620, and LD07-3395, respectively in field trials at Fayetteville, NC over a period of two years (2016–2017). The high yielding mutant FM6-847 was developed through ethyl methanesulfonate (EMS) mutagenesis at Southern Illinois University. The mutant line was derived from soybean cultivar Forrest and the dynamics of quality traits of the mutant was not assessed in North Carolina. Our analysis has revealed that the mean values of protein and amino acids in FM6-847 are approximately equal or higher than USDA checks. The majority of essential amino acid contents in soybean seeds are positively correlated each other and however, histidine content is negatively correlated with leucine ($r=-0.55$) and isoleucine ($r=-0.56$) contents. FM6-847 contains the highest amount of methionine (0.64%), tryptophan (0.43%), and sucrose (15.95%) but the lowest oil content (23.25%) compared to USDA checks. Almost all essential amino acid contents are negatively correlated with oil and sucrose contents, respectively but histidine content is positively correlated with oil content ($r=0.47$). Moreover, oil content is negatively correlated with sucrose ($r=-0.57$) and protein ($r=-0.46$) among the lines that were analyzed. Furthermore, histidine, phenylalanine, oil, and sucrose contents were significantly ($P<0.05$) impacted by environment (year by year) among these lines. Our findings will provide detailed information for soybean selection and QTL mapping on the target quality traits in future research.

PE0734: Legumes, Soybean, Common Bean, and related

Prediction of a Longitudinal Plant Trait Based on Marker Genotype and Environmental Data: An Application to Soybean Canopy Area Measured By UAV Remote Sensing

Yusuke Toda¹, Goshi Sasaki¹, Yoshihiro Ohmori¹, Yuji Yamasaki², Hirokazu Takahashi³, Hideki Takanashi¹, Mai Tsuda⁴, Hiromi Kajiya-Kanegae⁵, Hisashi Tsujimoto², Akito Kaga⁶, Mikio Nakazono³, Toru Fujiwara¹ and Hiroyoshi Iwata¹, (1)The University of Tokyo, Tokyo, Japan, (2)Arid Land Research Center, Tottori University, Tottori, Japan, (3)Nagoya university, Nagoya, Japan, (4)University of Tsukuba, Tsukuba, Japan, (5)Research Center for Agricultural Information Technology, NARO, Tsukuba, Japan, (6)National Institute of Agrobiological Sciences, Tsukuba, Japan

In plant breeding, unmanned aerial vehicles (UAVs) are expected to be useful devices for high-throughput phenotyping. Even for genomic selection, which enables the selection of genotypes without field testing, the development of high-throughput phenotyping systems is indispensable because it increases the size of training data and eventually the accuracy of genomic prediction. However, a method to predict longitudinal traits obtained by UAV remote sensing (UAV-RS) using genotype and environmental factors have not been developed. In this study, we developed a machine learning method to predict the growth pattern of the soybean vegetation area.

Two hundred genotypes of soybean germplasm were cultivated in Arid Land Research Center of Tottori University. The images of all plots were taken with UAV-RS for two years. The vegetation area of plant canopies on each observation day was estimated from image processing and statistical calibration.

We built a model to predict vegetation area based on both genotypic and environmental factors. The model predicted daily growth in the vegetation area based on marker genotype data and the environmental variables and vegetation area of the last three days. Random forest was chosen as a machine learning method for the modelling.

Prediction accuracy varied throughout the growth period, i.e., decreased as the growth proceeded due to variation in growth speed, and increased after the growth has saturated. Correlation coefficients of observed and predicted values at harvest ranged from 0.150 to 0.497. We are developing also a statistical method, which had a similar structure to a crop growth model, to predict the growth pattern of the soybean vegetation area.

PO0735: Legumes, Soybean, Common Bean, and related

Spatial Analysis and QTL Mapping to Resistance of Soybean to Stink BUG Complex

Emanoel Sanches Martins¹, Maiara de Oliveira¹, Patricia Braga¹ and José Baldin Pinheiro²,
(1)University of Sao Paulo - ESALQ, Piracicaba, Brazil, (2)University of Sao Paulo - ESALQ, Piracicaba, Sao Paulo, Brazil

The aim of this research was to fit a spatial model to adjust the natural dispersion of stink bug and to identify the QTLs that control the genetic architecture of the soybean resistance to the stink bug complex. The study was carried out with 256 RILs, derived from the cross between IAC-100 (resistant) and CD-215 (susceptible), being evaluated in the crop season 2017/2018, in Piracicaba. The experimental design was an alpha lattice 10x26 with three replicates. The evaluated traits were: graining period (GP), weight of a hundred seeds (WHS), grain yield (GY) and healthy seeds weight (HSW). To fit a Spatial Model we used the package SpATS including row and column information to adjust the model. The QTL mapping was performed through Composite Interval Mapping (CIM) considering the LOD thresholds of 1000 permutation test. We used a 1283 single-nucleotide polymorphisms (SNPs) to construct the linkage map and perform the QTL mapping. Including the field dispersion of stink bug in the model proved to be a good alternative for more accurate predicted values. The heritabilities were 0.65 to HSW, 0.86 to WHS, 0.57 to GY and 0.15 to GP. The QTL mapping allowed the identification of 7 QTLs, being 3 for HSW, 1 for WHS, 1 for GFP and 2 for GY. These QTLs are promising candidates to future studies as a genomic loci controlling the genetic architecture of stink bug complex resistance in soybean.

PE0736: Legumes, Soybean, Common Bean, and related

The Role of Syncytium Kinase Hub Genes in Soybean-Cyst Nematode Interactions

Sarbottam Piya, Miao Long, Bhoomi Patel, Hollis Rice, C. Neal Stewart Jr. and Tarek Hewezi,
Department of Plant Sciences, University of Tennessee, Knoxville, TN

Soybean cyst nematode (SCN) is a sedentary endoparasitic nematode that induces a specialized multinucleate feeding site, termed syncytium, in the root vascular tissues. The development of a functional syncytium involves considerable transduction signaling that orchestrates various cellular function and activity required for correct cell fate differentiation of this unique cell type. The goal of this study is to investigate the role of protein kinases, which constitute the core components of signal transduction pathways, in soybean-SCN interaction. We generated a comprehensive gene co-expression network and identified highly interconnected protein kinases that function as key nodes in signaling pathways. More than 200 kinase hub genes have been identified. Interestingly a significant number of these kinases were found to change expression in the SCN-induced syncytium, and hence were considered as syncytium kinase hub genes. The biological significance of the syncytium kinase hub genes was investigated using soybean transgenic hairy root system. Overexpression of inactive variants of a set of these kinase hubs significantly altered soybean susceptibility to SCN. These results provide strong evidence for the key roles that these kinases play in soybean-SCN interactions. Various functional assays are currently underway to reveal the mechanism through which these protein kinases mediate plant response to SCN.

PO0737: Legumes, Soybean, Common Bean, and related

Tilling By Exom Seq to Decipher Genes for Resistance to SCN

Mallory A. Cullen, Department of Plant Soil and Agricultural Systems, SIUC, Carbondale, IL

Soybean (*Glycine Max* (L.) Merr) is an important agricultural crop grown worldwide mainly for the production of oil and protein meal. The soybean cyst nematode (SCN), however, is the most damaging pest of soybeans causing over 1 billion US dollars in yield losses annually. The best disease management tool of SCN is through planting resistant lines. PI88788 type resistance accounts for 90 percent of SCN resistance used in commercial cultivars,

since most resistance is coming from the same source, the nematodes are overcoming this resistance mechanism creating a need for new types of resistance.

Targeting induced local lesions in the genome (TILLING) is a method for high throughput screening of a mutagenized population. TILLING can be utilized as a reverse genetics approach to identify functions of genes in different pathways. Our study focuses on using TILLING by Exom Seq for the identification of induced mutations within genes related to the SCN resistance pathway. A mutant library of 4000 mutagenized soybeans from a resistant cultivar Forrest was created, DNA from every mutant family extracted and arrayed in 2 dimensional pools, and probes designed targeting the exons of several genes within the soybean genome. After Illumina sequencing, the data revealed an abundance of single nucleotide polymorphisms (SNPs) between the mutants and wild type Forrest. Amongst these targeted genes were genes within the serine hydroxymethyltransferase gene family and pathogenesis related genes. The purpose of this study was to identify SNPs within these genes to better understand their roles in the SCN resistance pathway.

PE0738: Legumes, Soybean, Common Bean, and related

Molecular Cross Talk between Resistant and Defense Genes Promote Resistance to *Heterodera glycines*

Naoufal Lakhssassi, Southern Illinois University, Carbondale, IL, Khalid Meksem, Southern Illinois University at Carbondale, Carbondale, IL and Sarbottam Piya, Sadia Bekal, Vincent Colantonio, Shiming Liu, Zhou Zhou, Catherine Bergounioux, Long Miao, Jonas Meksem, Aicha Lakhssassi, Karen Jones, My Abdelmajid Kassem, Moussa Benhamed, Abdelhafid Bendahmane, Kris Lambert, Adnane Boualem, Tarek Hewez

Serine hydroxymethyltransferase (SHMT, E.C. 2.1.2.1) catalyzes the interconversion of serine/glycine as well as tetrahydrofolate (THF) and 5,10-methyleneTHF. Soluble NSF attachment proteins (α -SNAP) are conserved across animal and plant kingdoms and involved in vesicular trafficking, cytokinesis, and plasma membrane repair and stability. Pathogenesis related proteins (PRs) are widely present in plants and are induced following pathogen attack, elicitors, wounding, or stress, and are toxic to invading fungal pathogens. In the current study, we discovered the presence of a molecular interaction between the GmSHMT08 and GmSNAP18 proteins that is potentiated by the presence of a pathogenesis related protein GmPR10-08. GmPR10-08 was mapped to a novel quantitative trait locus for broad resistance to SCN using two different mapping populations. The GmPR10-08 was induced in response to SCN infections and, like GmSHMT08 and GmSNAP18, the PR10-08 was found to localize to the cytosol. Overexpression of GmPR10-08 decreased the number of SCN cysts to nearly 65% in transgenic soybean roots. A computational approach of structural homology modeling and docking algorithms reveals the predicted interaction sites between GmSNAP18, GmSHMT08, and GmPR10-08. Immunostaining and in-situ assays demonstrated that GmSNAP18 expression and localization cumulated at the plasma membrane and was specific to the root cells surrounding the nematode in SCN resistant soybean, but not in the SCN susceptible line, indicating the involvement of GmSNAP18 in molecular trafficking.

PO0739: Legumes, Soybean, Common Bean, and related

Using High-Throughput Sequencing to Characterize the Genetic and Genomic Architecture of Brown Stem Rot Resistance in Soybean

Chantal E. McCabe and Michelle A. Graham, USDA, Ames, IA

Breeding for pathogen resistance is an important objective to improve and protect yield. Brown stem rot (BSR), caused by the fungus *Phialophora gregata*, reduces soybean yield by as much as 38%. To date, three dominant BSR resistance genes have been identified: *Rbs1*, *Rbs2*, and *Rbs3*. However, the gene networks regulating defense responses to BSR remain unknown and the mapped location of all three loci is large and undefined. Identifying resistant germplasm by genotyping or phenotyping remains difficult due to complexities of soybean/*P. gregata* interactions. Therefore, the overarching goal of this postdoctoral research project was to characterize the genetic and genomic architecture of BSR resistance. To identify and characterize downstream defense genes, gene networks, and candidate resistance genes, RNA-seq of *P. gregata*-infected and mock-infected tissues of resistant (*Rbs1*, *Rbs2*, or *Rbs3*) and susceptible soybean genotypes was conducted. Preliminary analysis has revealed that one week after

infection, there is little overlap in differentially expressed genes between each resistant genotype. Further analyses will identify candidate genes for *Rbs1*, *Rbs2*, and *Rbs3* mediated resistance. Virus induced gene silencing (VIGS) has been used to characterize the genes and gene networks important in resistance. VIGS constructs were designed to target five clusters of receptor like proteins (RLPs) located within the three known *Rbs* loci. Silencing RLPs in resistant genotypes resulted in susceptibility to *P. gregata*, further validating their role in resistance. These results will increase the efficiency of identifying and developing cultivars with one or more BSR resistance genes, ultimately reducing yield loss due to BSR.

PE0740: Legumes, Soybean, Common Bean, and related

Mechanisms of Tolerance to Iron Deficiency in Soybeans

Bethany Q. Asmundson¹, Mohammed M Mira¹, Sylvie Renault¹, Robert D Hill² and Claudio Stasolla¹, (1)University of Manitoba, Winnipeg, MB, Canada, (2)University of Manitoba, Winnipeg, MB, Canada

Iron deficiency chlorosis negatively impacts soybean production worldwide. Iron is important in many physiological processes, including chlorophyll synthesis. Iron deficient plants exhibit interveinal yellowing and underdeveloped chloroplasts, which compromise overall growth, development, and final yield. These symptoms are commonly exacerbated by high soil pH and poorly drained soils. Strategies to cope with low iron availability include reduction of unusable Fe³⁺ into the usable form Fe²⁺, and to lower the accumulation of reactive oxygen species (ROS) responsible for oxidative damage. Changes in the antioxidant system, ascorbic acid availability, and reduced glutathione (GSH) facilitate reduction of Fe³⁺ to Fe²⁺ and mitigate ROS accumulation.

With the aim to elucidate the function of the antioxidant system during iron deficiency, twenty commercial soybean lines were screened for their ability to tolerate low iron levels. Based on chlorophyll content and signs of chlorosis, six lines (three tolerant and three susceptible to iron deficiency) were selected. Tolerant lines displayed increased activity of antioxidant enzymes generating ascorbic acid and GSH while susceptible lines exhibited increased activity of enzymes oxidizing ascorbic acid to monodehydroascorbic acid. Susceptible lines also displayed an increase in Phytoalbumin1 gene expression, a nitric oxide scavenging protein induced by diverse types of stress. Further studies, involving measurement of antioxidants, ROS, and the relative abundance of Fe³⁺ and Fe²⁺, will determine if the ability to accumulate higher levels of ascorbate and GSH represents an early strategy to mitigate iron deficiency in soybean. This information could be applied to predict plant performance to iron deficiency using simple and rapid physiological tests.

PO0741: Legumes, Soybean, Common Bean, and related

Quantitative Trait Loci Associated with Tolerance to Cold-Induced Seed Coat Discoloration in Soybean

Naoya Yamaguchi¹, Shizen Ohnishi², Seiji Hagihara², Dai Hirai³, Hideyuki Funatsuki⁴ and Mineo Senda⁵, (1)HRO Tokachi Agricultural Experiment Station, Kasai-gun, Hokkaido, Japan, (2)HRO Kitami Agricultural Experiment Station, Tokoro-gun, Hokkaido, Japan, (3)HRO Central Agricultural Experiment Station, Takikawa, Hokkaido, Japan, (4)NARO Institute of Crop Science, Tsukuba, Ibaraki, Japan, (5)Hirosaki University, Hirosaki, Aomori, Japan

Cold weather damages soybean [*Glycine max* (L.) Merr.] crops in high-latitude countries. Chilling temperatures negatively affect seed appearance by causing seed coat discoloration around the hilum region, which is called cold-induced discoloration (CD). In yellow soybean, seed coat pigmentation is inhibited by post-transcriptional gene silencing (PTGS) of chalcone synthase (*CHS*) genes. CD is caused by the suppression of *CHS* PTGS by chilling temperature. In yellow-seed coat with yellow-hilum cultivars, an inverted repeat of a *CHS* truncated sequence was suggested to be the *I* allele, and its double-stranded *CHS* RNA transcript is thought to induce *CHS* PTGS. An assay for CD tolerance using a phytotron was developed, and there were differences in CD tolerance among cultivars. The aims of this study were to identify quantitative trait loci (QTLs) associated with CD tolerance using a phytotron assay, and assess the effect of the major QTL under a cool field environment. Using the recombinant inbred lines between a CD-tolerant cultivar Toyoharuka and a CD-susceptible cultivar Toyomusume, the major QTL was detected in the proximal region of the *I* locus. Toyoharuka had a different allele at the *I* locus, which was designated *Ic* (inhibitor of CD). We confirmed that the *Ic* allele was highly effective using 27 cultivars and breeding lines

grown in the field where severe cold-weather damage occurred. This allele had no negative influence on the agronomic traits in the near-isogenic line. Our results suggest that marker-assisted selection for the *Ic* allele is effective for improving CD tolerance in breeding programs.

PE0742: Legumes, Soybean, Common Bean, and related

New Non-Homologous Duplicate Genes Controlling Short Petiole in Soybean

Kong Keke, Nanjing Agricultural University, Nanjing, China

Short petiole has potentials in soybean high-yield improvement and also performs special biological roles with insufficient studies. Here, a new recombinant line NG96-6755 were identified with short petiole. Compared with the normal plants, NG96-6755 has shorter internodes, smaller leaves, lower plant height. The results of paraffin section showed that the petiole cells of the mutant were smaller than those of the wild type. The segregation F_2 and $F_{2:3}$ generations of genetic populations fitted 15(long petiole):1(short) and 7(no segregation):8(segregation) theoretical ratios. It showed that short petiole of NG96-6755 was controlled by two pairs of recessive duplicate genes named as *sp1* and *sp2*. The two genes were roughly mapped to two genomic regions on chromosome 11 and chromosome 14 respectively using a next-generation sequencing-based bulked-segregant analysis approach. The regions were further delimited into 195kb and 183kb physical distances, and contained 18 and 17 candidate genes respectively. No gene of petiole length was reported in these non-homologous chromosome regions, so inferred that *sp1* and *sp2* might be new loci conferring soybean short petiole. The target gene was further determined by combining sequencing and qRT-PCR. This study provides theoretical basis and genetic materials for using short petioles to change plant architecture of soybean.

PO0743: Legumes, Soybean, Common Bean, and related

GmPGL1, a Thiamine Thiazole Synthase, Is Required for the Biosynthesis of Thiamine in Soybean

Xingxing Feng, Yaohua Zhang, Sunxin Yang, Kuanqiang Tang, Jiantian Leng, Jingjing Ma, Quan Wang and **Xianzhong Feng**, Northeast Institute of Geography and Agroecology, CAS, Changchun, China

Thiamine is an essential cofactor in several enzymatic reactions for all living organisms. Animals cannot synthesize thiamine and depends on their diet. Enhancing the content of thiamine is one of the most important goals of plant breeding to solve the thiamine deficiency associated with the low-thiamin staple crops. In this study, a *Glycine max* *pale green leaf 1* (*Gmpgl1*) mutant was isolated from the EMS mutagenized population of soybean cultivar, Williams 82. Map-based cloning of the *GmPGL1* locus revealed a single nucleotide deletion at the 292th nucleotide residue of the first exon of *Glyma.10g251500* gene in *Gmpgl1* mutant plant, encoding a thiamine thiazole synthase. Total thiamine contents decreased in both seedlings and seeds of the *Gmpgl1* mutant. Exogenous application of thiazole restored the pale green leaf phenotype of the mutant. The deficiency of thiamine in *Gmpgl1* mutant led to reduced activities of the pyruvate dehydrogenase (PDH) and pyruvate decarboxylase (PDC), and decreased contents of six amino acids as compared to that in the wild type plants. These results revealed that *GmPGL1* played an essential role in thiamine thiazole biosynthesis.

PE0744: Legumes, Soybean, Common Bean, and related

QTL Mapping for Kernel Matter Accumulation in Peanut (*Arachis hypogaea* L.)

Xiaoxia Zhang, Hui Yang, Xiurong Zhang, Yuying Lv, Kun Zhang, Yongshan Wan and **Fengzhen Liu**, State Key Laboratory of Crop Biology, Shandong Key Laboratory of Crop Biology, Shandong Agricultural University, Tai'an, China

The matter accumulation is one of the key factors to affect the fullness degree of kernel, and directly affect the yield and quality of peanut. Genetic research on the traits related to matter accumulation of kernel is essential to improvement of peanut variety.

In this study, the shrinking kernel mutant line (05D677) and the wild type line (05D610) were used as parents to construct F_2 population, which composed of 214 individuals. The ratio of normal plants to shrinking kernel plants was 2.45:1 in F_2 population, and the kernel shrinkage trait is controlled by a recessive gene according to the chi-

square test. We completed a whole-genome re-sequencing of two DNA bulks (shrinkage pool and normal pool) generated from plants in F₂ population, and DNA from two parent lines (05D677 and 05D610), respectively. A genome-wide analysis of single nucleotide polymorphisms (SNPs) resulted in the detection of 3316 SNPs between female parent 05D677 and male parent 05D610. Three genomic regions significantly associated with kernel matter accumulation trait were mapped: Arah1.09:73103254 bp~76940606 bp, Arah1.15:151664075 bp~1594450926 bp and Arah1.15:61236917 bp~63969695 bp. *Arah1.M8GUZX*, *Arah1.LIRE4J* and *Arah1.QV02Z8* in the region of Arah1.15:151664075 bp~1594450926 bp were predicted as potential candidate genes.

Keywords: Peanut (*Arachis hypogaea* L.); Matter accumulation; Kernel shriveling; Bulk segregant analysis; QTL mapping

PO0745: Legumes, Soybean, Common Bean, and related

Identification and Association of Differentially Expressed Disease Resistance (R) Gene Candidates Involved in Leaf Spot Resistance in Peanut (*Arachis hypogaea* L.)

Phat M. Dang¹, Marshall C. Lamb¹ and Charles Y. Chen², (1)USDA/ARS/National Peanut Research Laboratory, Dawson, GA, (2)Department of Crop, Soil and Environmental Sciences, Auburn University, Auburn, AL

Early (ELS) and late leaf spot (LLS) are fungal diseases that cause severe defoliation and can significantly limit peanut production in the United States and around the world. Breeding for high resistance in peanut has been challenging due to strong genotype by environment interaction. These complex traits are controlled by many major and minor quantitative trait loci (QTLs) involving many genes. Furthermore, chromosome locations of the QTLs varied depending on populations or methods utilized for analysis. The goals of this research were to 1) identify candidate genes for leaf spot resistance, and 2) to associate gene-expression to resistance. Candidate genes identified included TMV resistance protein N-like, PTII-like tyrosine-protein kinase, pto-interacting protein, cysteine-rich receptor-like protein kinase, and phyto-sulfokine receptor-like. From qPCR analysis utilizing leaf tissues from leaf spot susceptible and tolerant genotypes in field evaluations, several R genes on chromosomes 5 and 8 were differentially regulated and strongly associated with leaf spot resistance. Candidate R genes located on other chromosomes will be evaluated as well. Gene-expression levels and patterns will be associated with peanut genotypes with leaf spot resistance. This research will facilitate the development of peanut varieties with high leaf spot resistance.

PE0746: Legumes, Soybean, Common Bean, and related

Validation of the Utility of GWAS-Derived Markers for Tolerance to Water Deficit in a Segregating Peanut Breeding Population

Mark D. Burow¹, Jennifer Chagoya², Roshan Kulkarni³, Vikas Belamkar⁴, Ratan Chopra⁵, Michael R. Baring², Charles Simpson⁶, John Cason⁶, Paxton Payton⁷ and James Mahan⁷, (1)Texas A&M AgriLife Research, Lubbock, and Department of Plant and Soil Science, Texas Tech University, Lubbock, TX, (2)Texas A&M AgriLife Research, Lubbock, TX, (3)Iowa State University, Ames, IA, (4)Department of Agronomy and Horticulture, University of Nebraska-Lincoln, Lincoln, NE, (5)University of Minnesota, St. Paul, MN, (6)Texas A&M AgriLife Research, Stephenville, TX, (7)USDA-ARS-CSRL, Lubbock, TX

The U.S. peanut minicore collection was found previously to have some accessions that outyielded commercial cultivars under water deficit stress. SSR markers for yield and associated traits under water deficit were identified by association mapping. A population was developed to combine tolerance to water deficit from a minicore accession, with resistance to root-knot nematodes and the seed high oleic fatty acid content derived from a breeding line. Initial selection and testing were performed in the F₂ generation. Accessions were advanced and breeding lines were grown as replicated trials from 2015 to 2019. Progeny of accessions selected based on marker scores in the F₂ generation outyielded the other accessions by 20%. Several accessions yielded well repeatedly under water deficit compared to commercial varieties. Large differences in rankings between irrigated and water deficit conditions suggested that certain accessions possess tolerance to water deficit, as opposed to high yield potential in general. However, there were no markers for grade, and the low shelling among accessions is thought to be a result of use of the unadapted

minicore material as parent. It is proposed to use marker-assisted backcrossing to select for yield and higher grade as well as other key traits.

PO0747: Legumes, Soybean, Common Bean, and related

Identifying Genomic Regions Associated with Seed Quality and Germination Traits in Mini-Core Peanut Population

Jinesh D. Patel, Crop Soil and Environmental Science, Athens, AL and Charles Y. Chen, Department of Crop, Soil and Environmental Sciences, Auburn University, Auburn, AL

Traits like seed size, seed weight, small shelled kernels, kernel weight, hull weight determines quality of peanut seed. Numerous genomic region and gene network are regulating such trait. Identifying such genomic regions will help breeders to develop molecular marker for MAS (Marker Assisted Selection) breeding. A total of 120 accessions from U.S. peanut mini core collection were evaluated for seed quality traits. These accessions were also genotyped using 58K SNP array and we were able to identify 17K high quality SNP for this association panel. Base on results, we observed significant variation for seed quality traits in different accessions and different botanical varieties. Through Genome Wide Association Study (GWAS), we were able to identify multiple regions associated to small shelled kernels, seed weight, kernel weight, hull weight. For instance, marker AX-176823847 (Chr 15), AX-176794068 (Chr 12), AX-177638040 (Chr 10), AX-176794068 (Chr 11) and AX-147216060 (Chr 03) were strongly associated with seed size, small shelled kernel, seed weight, kernel weight and hull weight. Areas surrounding these markers were scrutinized for candidate genes associated with these traits. Multiple genes were identified in the regions that might have important role during seed development. In summary, our work will provide markers that could be incorporated in breeding program to accelerate selection process for seed quality and explore the possibility of function of candidate gene to understand the complex genetic network that governs seed quality.

PE0748: Legumes, Soybean, Common Bean, and related

Cowpea Genome Information Resources

Timothy J. Close¹, Stefano Lonardi², Shengqiang Shu³, Maria Munoz-Amatriain⁴, Sassoum Lo⁵, Pei Xu⁶, Andrew D. Farmer⁷, Samuel A. Hokin⁷, Qihua Liang², Jodi L. Humann⁸, Steven B. Cannon⁹, Joseph W. Carlson¹⁰, James Crabb¹¹, Patrick A. Davidson¹⁰, David M. Goodstein³, Yi-Ning Guo¹², Richard D. Hayes¹³, Dorrie Main¹⁴, Jeremy Phillips³, Philip A. Roberts¹ and Xinyi Wu¹⁵, (1)University of California, Riverside, Riverside, CA, (2)Department of Computer Science & Engineering, University of California, Riverside, CA, (3)DOE Joint Genome Institute, Walnut Creek, CA, (4)Department of Soil and Crop Sciences, (5)University of California Davis, (6)China Jiliang University, China, (7)National Center for Genome Resources (NCGR), Santa Fe, NM, (8)Washington State University, Pullman, WA, (9)USDA–Agricultural Research Service, Corn Insects and Crop Genetics Research Unit, Ames, IA, (10)DOE Joint Genome Institute, Berkeley, CA, (11)Department of Horticulture, Washington State University, (12)Department of Botany & Plant Sciences, University of California - Riverside, Riverside, CA, (13)US Department of Energy Joint Genome Institute, Berkeley, CA, (14)Department of Horticulture, Washington State University, Pullman, WA, (15)Zhejiang Academy of Agricultural Sciences, China

Cowpea (*Vigna unguiculata* [L.] Walp.) is a diploid warm-season legume, also known as black-eyed pea, among many other names. Cowpea is relevant mainly as a grain legume in the USA, Europe and Latin America, and as a fresh vegetable (longbean) in China and elsewhere in Asia. It is also of major importance as food and fodder in sub-Saharan Africa. The Global Food Security Reauthorization Act of 2018 (H. R. 5129) identified twelve target countries, several of which are African nations that rely on cowpea. Here we summarize progress on cowpea pangenome sequencing and annotation, focused so far on cultivated accessions including one representative of each of five sub-populations (IT97K-499-35 from the IITA breeding program in Nigeria, CB5-2 bred in California, Suvita2 as a landrace from Burkina Faso, Sanzi as a landrace from Ghana, and UCR779 as a landrace from Botswana) and two longbean accessions from the Asian sub-population (elite TZ30, and landrace ZN016). Several online portals provide various levels of access to information (e.g. genes, gene expression, variants) related to the IT97K-499-35 genome, which was the first well-sequenced accession among these seven (Lonardi et al., 2019, Plant Journal 98:767-782), and to the others that have been sequenced and annotated more recently. These portals are

Phytozome (phytozome.net), the Legume Information System including Legume Mine (legumeinfo.org), NCBI Genome (ncbi.nlm.nih.gov/genome), and the Pulse Crop Database (pulsedb.org). Work is underway to provide interactive visual representations of multiply aligned genome sequences to provide facile access to features of the cowpea pangenome. Several US federal funding sources have supported or are currently supporting the development of this information and the teams that are now building these cowpea genome resources: NSF, USAID, USDA, DOE and NIH. The National Natural Science Foundation of China and the National Ten-Thousand Talents Program of China have supported the longbean sequencing. Additional sources of support are also noted.

PO0749: Legumes, Soybean, Common Bean, and related

Genomic Predicted Breeding Values of Cowpea Lines Grown in Low Phosphorous Soils

Saba Baba Mohammed¹, Muhammad F. Ishiyaku², Muhammad L. Umar², P. B. Tongoona³ and Daniel D Dzidzienyo³, (1)Ahmadu Bello University, Zaria, Nigeria, (2)Institute for Agricultural Research Ahmadu Bello University, Zaria, Nigeria, (3)West Africa Centre for Crop Improvement, Accra, Ghana

The grain yield of cowpea is limited by low soil phosphorus (P) in major producing areas of the crop in sub-Saharan Africa. Genetic improvement of the crop for phosphorus use efficiency is crucial for developing and deploying varieties that fit into soils with sub-optimal fertility. Genomic selection (GS) is believed to increase rates of genetic gain and reduce breeding cycles for quantitative traits, but few trials implementing GS have been conducted for cowpea. Therefore, our objectives were to use phenotype and genotype data acquired through genotyping by sequencing to predict breeding values of 400 cowpea lines evaluated under low soil P conditions to select parents with favourable alleles and to shorten the cycle of selections to develop improved P efficient varieties. Prediction models were trained with phenotype from two 2 year experiments, with each year replicated twice. For GS, markers were DarTSeq generated using the genotyping-by-sequencing platform. Parents with favourable alleles for grain yield in low P field environments were identified and are being crossed to generate F1 progenies. Further work to develop efficient GS implementation strategies in cowpea is ongoing.

PE0750: Legumes, Soybean, Common Bean, and related

Leveraging Legume Genomic Resources to Identify Genes Associated with Non-Seed Shattering in Hairy Vetch (*Vicia villosa* Roth)

Tim D. Hernandez¹, Yanina Alarcon¹, Lisa Kissing Kucek², Nick Krom¹, Christy M. Motes¹, Shahjahan Ali¹, Heathcliffe Riday² and Maria J. Monteros¹, (1)Noble Research Institute, Ardmore, OK, (2)US Dairy Forage Research Center (USDA-ARS), Madison, WI

Hairy vetch (*Vicia villosa* Roth) is a versatile annual legume with utility as both a forage and cover crop. It has the capacity for biological nitrogen fixation, serves as a ground cover to protect from soil erosion and improves soil structure. Seed shattering is the natural ability of plants to disperse seed and continue with the next generation. However, seed shattering is undesirable in agricultural systems as it presents challenges for seed harvest that leads to seed yield losses. Selection against seed shattering has been successful as part of the domestication process, but non-seed shattering is still a target of improvement in hairy vetch and other legume breeding programs. The objective of this project is to identify genetic components associated with non-seed shattering in hairy vetch leveraging existing resources in legumes and complemented with association studies with genotypes contrasting for seed shattering in the field. Genotyping-by-sequencing (GBS) enabled characterization of sequence variants in hairy vetch accessions with a range of seed shattering phenotypes. Molecular pathways and regulatory elements involved in pod dehiscence including transcription factors catalogued in other species served as templates to capture sequence variants in hairy vetch. Additional exploration of differential gene expression in common vetch resulted in additional gene targets to complement the genome-wide SNP discovery efforts to explain the genetic variation in the seed shattering phenotype. Deployment of comparative genomics approaches combined with transcriptome mining and genome-wide SNP discovery efforts can facilitate understanding of the genetic mechanisms underlying non-seed shattering in hairy vetch. The genomic tools developed can serve to launch future genome-wide association studies and fine mapping to inform selection decisions in breeding programs aimed at developing cultivars with seed retention capabilities.

PO0751: Legumes, Soybean, Common Bean, and related

Transcript Profiling of Hairy Vetch (*Vicia villosa*)

Shahjahan Ali¹, Sarah A. Krogman², Kimberly Cooper², Lynne Jacobs², Nick Krom¹, Perdeep K. Mehta², Michael Trammell², Suresh Bhamidimarri³, Twain Butler² and Maria J. Monteros¹, (1)Noble Research Institute, Ardmore, OK, (2)Noble Research Institute, LLC, Ardmore, OK, (3)Corteva Agriscience, Connell, WA

Hairy vetch, a diploid annual legume species, has a robust growth habit, high biomass yield and winter hardiness characteristics. The capacity for symbiotic nitrogen fixation and ground cover contribute the use of hairy vetch as both a forage and cover crop legume. Seed hardness affects the ability of the seed to imbibe water and germinate under certain conditions. Hard-seeded cultivars are valuable as forages, while soft-seeded cultivars provide advantages for their use as a cover crop. The objectives of this study are to understand the genetic mechanisms associated with seed hardness in hairy vetch through a combination of approaches including transcript profiling and mapping approaches. RNA from leaves, flowers, immature pods, seed coats and cotyledons from plants derived from the AU Merit cultivar and contrasting for soft vs. hard seeds was extracted. Illumina's HiSeq4000 platform was used to generate between 31.22 to 79.18 Gb RNA sequence data per tissue sample. Contig assembly and mapping of the contigs against the *Medicago truncatula* (V4.0) genome resulted in identification of 72,466 gene transcripts across all tissues. Differential gene expression analysis between hard vs. soft-seeded types, enabled identification of genes up- and down-regulated in seed coats compared to other plant tissues. Key candidate genes with a potential role in seed hardness were further refined using known genes involved in seed hardness in other species to query the hairy vetch transcriptome data. Identification of genetic determinants of seed hardness in hairy vetch can facilitate the development of improved cultivars with desirable seed characteristics for dual-purpose use.

PE0752: Legumes, Soybean, Common Bean, and related

The Evolutionary History of *Phaseolus vulgaris* As Revealed By Chloroplast Genome

Giulia Frascarelli¹, Elena Bitocchi¹, Teresa Rosa Galise², Nunzio D'Agostino², Salvatore Cozzolino³ and Roberto Papa¹, (1)Department of Agricultural, Food and Environmental Sciences, Polytechnic University of Marche, Ancona, Italy, (2)Department of Agricultural Sciences - Division of plant genetics and biotechnology University of Naples Federico II, Napoli, Italy, (3)Department of Biology University of Naples Federico II, Napoli, Italy

Knowledge about the origin, evolution and diffusion of crop species is a crucial aspect for their appropriate use and conservation. *Phaseolus vulgaris* has a unique evolutionary history, with the wild form originated in Mesoamerica and subsequently introduced in South America, leading to the formation of two South American wild gene pools in North Peru and Ecuador and in South Andes. However, a debate is still open about common bean origin. Indeed, recent studies proposed a new hypothesis on the origin of common bean, the so-called "*Pseudovulgaris*" hypothesis, which indicates that the formation of the North Peru and Ecuador gene pool occurred much earlier than the formation of *P. vulgaris* species and, thus of the diversification of Mesoamerican and Andean gene pools. This suggests that the North Peru-Ecuador population represents a different species, named *P. pseudovulgaris* and it shared a common ancestor with the Mesoamerican and Andean groups, that remains to be discovered or has become extinct. The aim of this work is to clarify the phylogeny of *P. vulgaris*. A large sample that represents the entire geographical distribution of the wild forms of this species was investigated by analyzing chloroplast genome diversity inferred from WGS (Whole Genome Sequencing) data. Thirty-nine *de novo* chloroplast genomes were assembled. Our results corroborate monophyletic and Mesoamerican origin of common bean, not supporting the recently proposed *Pseudovulgaris* hypothesis.

PO0753: Legumes, Soybean, Common Bean, and related

Fine Mapping the Rust Resistance Gene *Ur-6* in Common Dry Bean

Peter E. Beerbower¹, Atena Oladzadabbasabadi¹, Rian Lee² and Phillip McClean³, (1)North Dakota State University, Genomics and Bioinformatics Program, Fargo, ND, (2)Dept of Plant Sciences, North Dakota State University, Fargo, ND, (3)North Dakota State University, Fargo, ND

Bean rust, a disease caused by *Uromyces appendiculatus*, is a serious threat to the farming of common dry bean (*Phaseolus vulgaris* L.) that completes its entire life cycle on common bean. *Ur-6* confers resistance to races of *U.*

appendiculatus that overcome other resistance genes and is present within the Andean gene pool, which is more commonly farmed in South America and Africa. In this study we attempted to finely map *Ur-6* to the common bean genome. To this end, we have screened the Middle American Diversity Panel with *U. appendiculatus* race 15-3 (47), which identifies *Ur-6*. Efficient Mixed-Model Association (EMMA) was carried out to identify 130 SNPs highly associated with trait race 15-3 (47) resistance ($p \leq 8.89E-05$). InDel markers were screened with association with trait race 15-3 (47) resistance in an F_2 population generated from Golden Gate Wax (*Ur-6*) and UI-114 (no resistance genes). Preliminary analysis has identified a strong association of trait race 15-3 (47) resistance with a gene cluster on the proximal end of Pv07.

PE0754: Legumes, Soybean, Common Bean, and related

Genome-Wide Association Studies and ROS Production during the Response of Common Bean to *Meloidogyne incognita* Infection

Willian Giordani¹, Augusto L. Diniz², Henrique C. Gama¹, Alisson F. Chiorato³, João P. R. Marques¹, Maria Lucia C. Vieira¹ and Antonio A. F. Garcia¹, (1)University of São Paulo (ESALQ/USP), Piracicaba, Brazil, (2)Department of Biochemistry, Institute of Chemistry, University of São Paulo, São Paulo, Brazil, (3)Grain and Fiber Center, The Agronomic Institute of Campinas (IAC), Campinas, Brazil

The southern root-knot nematode (RKN; *Meloidogyne incognita*) is one of the main limiting factors for common bean production, with losses up to 90%. The main challenges to the development of resistant cultivars are the phenotyping approaches that are time-consuming, laborious and often involve inaccuracies. Therefore, identifying resistance genes and markers to assist selection is essential to improve breeding efficiency. Furthermore, although common bean studies are still incipient, in analyses on other species there are reports that reactive oxygen species (ROS) production plays an important role in the defense against RKN. In this study we aimed to identify, via genome-wide association studies (GWAS), genomic regions associated with the common bean response to RKN and verify the early production of ROS in contrasting genotypes. We used an association panel of 175 genotypes, which were evaluated for root-galling and number of egg-masses by a high-throughput hydroponic-type assay. The plants were grown in growth pouches with nutritive solution and inoculated with 1,500 second-stage juveniles. Thirty days post-inoculation, egg-masses were counted and root-galling indexed. The three most resistant and susceptible genotypes were also evaluated for ROS production 12 h post-inoculation. Genotyping by sequencing (GBS) permitted the identification of 10,362 SNPs that were used for GWAS, four of which were associated with root-galling (Pv1, 2, 5, 10) and four with egg-masses (Pv6, 7, 8, 11). Regarding ROS, our results indicate stronger production in resistant genotypes compared to susceptible genotypes, suggesting that these molecules also act during the common bean response to RKN.

PO0755: Legumes, Soybean, Common Bean, and related

Root Nodulation Requires Multiple 3-Hydroxy-3-Methylglutaryl Coenzyme A Reductases in *Medicago truncatula*

Zachary P. Keyser, University of Wisconsin-Madison, Madison, WI

Symbiotic associations between legumes and rhizobia are controlled by a dedicated signaling pathway in the roots of legume plants. Genetic and biochemical studies in the model legume *Medicago truncatula* led to the identification of receptors (MtLYK3 and MtNFP) for microbial signals, a co-receptor (MtNORK) at the plasma membrane level, as well as downstream components involved in nuclear calcium spiking, which ultimately regulates the expression of genes associated with symbiosis. We found a 3-Hydroxy-3-Methylglutaryl Coenzyme A Reductase (MtHMGR1) as an interactor of MtNORK. Additionally, mevalonate, the product of HMGR activity, is sufficient to trigger nuclear calcium spiking even in the absence of MtNORK. MtHMGR1 is, therefore, an excellent candidate to connect perception events at the plasma membrane level to nuclear ones such as calcium spiking and gene expression. We will present new data indicating that MtHMGR1 also interacts with the MtLYK3 and MtNFP receptors. Also, RNAi silencing of HMGR1 expression decreases calcium spiking, symbiotic gene expression, and nodulation drastically. However, when the fully sequenced genome of *M. truncatula* was released, we identified several new homologs of MtHMGR1 that were likely affected by our RNAi construct. To determine which specific HMGRs are involved in symbiosis, we used *Tnt1*-insertion lines, knocking-out the expression of individual HMGRs. Insertions in MtHMGR1 and MtHMGR2c decreases the ability of nodules to support nitrogen fixation but not nodule number,

indicating that multiple HMGRs are involved in the nodulation process with a likelihood of some functional redundancy.

PE0756: Legumes, Soybean, Common Bean, and related

Beyond *Pisum sativum* Genome

Jonathan Kreplak, Gregoire Aubert, Nadim Tayeh and Judith Burstin, INRAE, UMR1347 Agroécologie, Dijon, France

Pea (*Pisum sativum* L.) is the second most important grain legume in the world after common bean and Gregor Mendel's original genetic model. With a genome size of 4.3Gb and a large amount of transposable element, *Pisum sativum* was lagging behind other important legume crops for genomics development. We reported the first annotated chromosome-level reference genome assembly for pea (J. Kreplak et al 2019). Compared to other sequenced Leguminosae genomes, the pea genome shows intense gene dynamics, most likely associated with genome size expansion when the Fabeae diverged from its sister tribes. Other species like *Vicia faba*, with genome larger than pea, are still not sequenced and the sequence of pea is a valuable resource to have insight on gene conservation and tackle translational genomics approaches.

In this poster, using mainly transcriptomics data from lentil, grass pea and faba bean, we provide a first overview of gene content comparison and some possible applications.

PO0757: Legumes, Soybean, Common Bean, and related

Integrating Genetics, Genomics and Transcriptomics to Understand Flowering Time Responses in Lentil Under Different Light Quality Environments

Hai Ying Yuan¹, Carolyn T. Caron¹, Larissa Ramsay¹, James L. Weller², Bert Vandenberg¹ and Kirstin Bett¹, (1)University of Saskatchewan, Saskatoon, SK, Canada, (2)School of Natural Sciences, University of Tasmania, Hobart, Australia

Appropriate responses to both light quantity and quality are necessary for the adaptation of plants to specific environments. Sub-optimal light environments like low Red/Far-red related light quality induces a collective photomorphogenetic response in plants. In general, far-red light enriched under high-density cultivation or weedy environments results in a low red to far-red ratio, which leads to earlier flowering in many species. In our study, we used an interspecific RIL population (*L. culinaris* X *L. orientalis*), based on parents showing contrasting sensitivity to changes in light quality, to examine the flowering responses towards different light quality environments in lentil. A high-density SNP linkage map and phenotyping from controlled growth chamber trials were used to identify QTLs. Transcriptomic analysis using RNAseq, together with the annotated Lentil genome revealed candidate genes controlling flowering time responses under conditions of differing light qualities. Combining these different approaches, we seek to have better understanding towards the flowering-time adaptation of *Lens* spp. to suboptimal light environments. Increased understanding of light responses will help improve our ability to develop cultivars that have better adaptation to variable light environments.

PE0758: Legumes, Soybean, Common Bean, and related

Genome-Wide Association Study (GWAS) of Salinity Tolerance in Australian Lentil Germplasm

Ruwani Dissanayake¹, Noel O.I. Cogan^{2,3}, Garry Rosewarne⁴, Kevin Smith^{1,5} and Sukhjiwan Kaur⁶, (1)University of Melbourne, Melbourne, VIC, Australia, (2)Agriculture Victoria, AgriBio, Centre for AgriBioscience, Bundoora, VIC, Australia, (3)School of Applied Systems Biology, La Trobe University, Bundoora, VIC, Australia, (4)Agriculture Victoria, Grains Innovation Park, Horsham, VIC, Australia, (5)Agriculture Victoria, Hamilton Centre, Hamilton, VIC, Australia, (6)Agriculture Victoria, Bundoora, VIC, Australia

Soil salinity is identified as one of the major abiotic stress factors limiting crop productivity across the world. In Australia, many agricultural soils are affected by high levels of salt that can result in reduced yield due to poor plant growth as well as reduction in water and nutrient uptake. In comparison to other grain crops including wheat and

barley, lentil is more salt sensitive crop with more than 90% potential yield loss at EC = 3 dS/m. A diversity panel of c. 276 advanced breeding lines was assessed to perform a genome-wide association study (GWAS), to identify genomic regions conferring salinity tolerance in lentil. The genotyping was performed using a newly developed and optimized targeted genotyping-by-sequencing (tGBS) method and c. 56,349 genome wide single nucleotides polymorphic (SNP) markers were identified. Phenotyping was performed using a randomized complete block design with four replicates and stress tolerance was evaluated based on visual salinity symptoms and shoot dry weight reduction. GWAS for salinity tolerance was conducted using TASSEL, through a mixed linear model based on kinship. Marker-trait association were observed on Chromosome 2 and a total of 12 significant SNP markers were identified (False Discovery Rate; FDR>0.05). Comparison of these marker positions to gene-finding format (GFF) file from CDC Redberry genome (v2.0) identified salt responsive family proteins as candidate genes for further validation. The results of our study will provide information on new sources of tolerance present in Australian lentil germplasm and mechanism behind salt tolerance in lentil.

PO0759: Legumes, Soybean, Common Bean, and related

Characterizing Nodulation Gene Mutations in Chickpea (*Cicer arietinum*)

Daniel Frailey and Thomas M. Davis, University of New Hampshire, Durham, NH

In addition to being an important food source for much of the world's population, legumes are also rotation crops because their symbiotic relationship with *Rhizobium* bacteria increases soil nitrogen. Nodulation is the formation of specialized organs called root nodules, and is essential for this symbiosis. *Rhizobium* are capable of fixing atmospheric nitrogen and providing it to the plant in a usable form. Two radiation-induced, mutant lines of chickpea, PM233 and PM405, are deficient in nodule formation and/or function. We used Illumina reads aligned to the Desi uwa-V3.0 chickpea reference genome to identify a candidate gene and mutation site for the PM405 mutation: the homologue of *Medicago truncatula* nodulation gene *NSP1* located on chickpea pseudochromosome 2. The mutant allele co-segregated with the PM405 phenotype in a progeny population providing further evidence of its responsibility for the defective nodulation phenotype. Based on marker-trait association, the gene responsible for the PM233 phenotype was identified by other investigators as *CaNSP2*, the chickpea homologue of *M. truncatula* nodulation gene *NSP2*. But the structure of the PM233 mutation resulting in loss of function has never been characterized. Alignment of Illumina reads from PM233 to the Desi uwa-V3.0 reference genome located the candidate gene site, but could not resolve one boundary of an evident deletional event. We then used Illumina and Nanopore sequencing data from wild type desi line ICC640 to construct contigs corresponding to chickpea pseudochromosome 5. Alignment of Illumina and Nanopore reads to one of these contigs enabled clear definition of the putatively causal deletion.

PE0760: Legumes, Soybean, Common Bean, and related

Crimson Clover (*Trifolium incarnatum*) Cultivar Development Using Marker Assisted Selection Methods for the Southern Plains

Sarah A. Krogman¹, Shahjahan Ali² and Michael Trammell¹, (1)Noble Research Institute, LLC, Ardmore, OK, (2)Noble Research Institute, Ardmore, OK

Winter annual legumes are a sustainable alternative to the current continuous wheat rotation system used in the Southern Plains. It can be used as a forage, cover crop or inter-crop for silvopastoral systems. The addition of winter legumes to agro-eco systems increases soil fertility by adding nitrogen through symbiotic nitrogen fixation system by root nodule bacteria. It also breaks pest and disease cycles when used in rotations. When used in forage systems, they can increase the nutritional plane of grazing animals and reduce the amount of nitrogen fertilizer application. Crimson clover (*Trifolium incarnatum*) is an excellent legume candidate to incorporate within these systems as it exhibits many desirable traits such as, early fall establishment, winter hardiness and high biomass yield. In order to develop new cultivars of crimson clover adapted for the Southern Plains, 48 accessions, acquired from the Germplasm Resources Information Network (GRIN) are being evaluated in replicated trials in multiple locations and years. Leaf tissues were collected from each accession with several genotypic representatives per line. Genomic DNA will be isolated from these leaf samples and single nucleotide polymorphism (SNP) markers will be generated from a GBS (Genotyping-by-sequencing) method. Marker-trait association studies will be performed in Genome Wide Association Study (GWAS) method. Both phenotypic and Marker Assisted Selection (MAS) will be employed to select superior genotypes for cultivar development.

PO0761: Legumes, Soybean, Common Bean, and related

***Medicago sativa* L (Alfalfa) Employs Molecular Physiological Strategies to Regulate Drought Response Through the miR156-SPL13-Dfr Module**

Abdelali Hannoufa^{1,2}, Biruk Feyissa² and Susanne Kohalmi², (1)Agriculture and Agri-Food Canada, London Research and Development Centre, London, ON, Canada, (2)University of Western Ontario, London, ON, Canada

Developing alfalfa cultivars that can withstand drought is critical for the sustainable production of this forage crop. miR156 is highly conserved in plants, where it functions by silencing a group of *SQUAMOSA-PROMOTER BINDING PROTEIN-LIKE (SPL)* transcription factors. We previously showed that miR156 overexpression in alfalfa improves drought tolerance. In this study, we used three alfalfa genotypes with different levels of miR156 overexpression and three genotypes with reduced *SPL13* expression to study drought tolerance strategies by miR156 at the phenotypic, physiological, metabolic and molecular levels. These analyses on stems, roots and leaves of the used genotypes revealed a coordinated response to drought mediated by *SPL13* silencing. Low to moderate levels of miR156 improved drought tolerance in alfalfa by silencing *SPL13*, increasing accumulation of stress mitigating metabolites such as proline, gamma-aminobutyric-acid, anthocyanins and other flavonoids, as well as enhancing photosynthetic assimilation rate, Fv/Fm ratio and root development. Moreover, transcription of genes involved in secondary metabolite synthesis and photosynthesis were increased significantly in the moderate miR156 over-expressors. We also demonstrate that the *SPL13* protein binds to the promoter region of *DIHYDROFLAVONOL-4-REDUCTASE* gene *in vivo* to regulate its expression level. We conclude that a moderate increase in miR156 levels (0.5 to 1.5) is sufficient to enhance drought resilience in alfalfa, but higher miR156 overexpression may result in drought susceptibility.

PE0762: Legumes, Soybean, Common Bean, and related

Genomic Approach for Enhancing Abiotic Stress Resilience in Alfalfa (*Medicago sativa* L.)

Long-Xi Yu¹, Cesar Medina², Sen Lin², Michael Peel³, Meijing Liang⁴ and Zhiwu Zhang⁵, (1)USDA-ARS Plant Germplasm Introduction and Testing Research, PROSSER, WA, (2)USDA-ARS, Prosser, WA, (3)USDA, Logan, UT, (4)Dept. of Crop and Soil Science, Washington State University, Pullman, WA, (5)Washington State University, Pullman, WA

Alfalfa is a worldwide forage legume and is “queen of forage” due to its high nutritional value. However, the production of alfalfa is challenged by adverse environmental factors such as drought and high salinity. Developing resistance alfalfa is an important breeding target for enhancing alfalfa productivity in arid and semi-arid regions. Since drought and salinity occur simultaneously in these regions. To understand the genetic base of abiotic stress resilience, two alfalfa populations were used for evaluating drought and salt resistance in the field. Genome-wide association studies with genotyping by sequencing were used for mapping resistance loci. Twenty-eight markers were associated with yield under drought. Most of the markers were identified across different harvest periods under water deficit, although different levels of significance were found among the three harvests. The loci associated with biomass yield under water deficit located throughout all chromosomes in the alfalfa genome agreed with previous reports. Our results suggest that biomass yield under drought may involve a different mechanism compared to that of non-stress. BLAST searches of the flanking sequences of the associated loci against DNA databases revealed several stress-responsive genes linked to the drought resistance loci, including leucine-rich repeat receptor-like kinase, B3 DNA-binding domain protein, translation initiation factor IF2 and phospholipase-like protein. Marker-trait association identified a total of 42 markers significantly associated with five traits associated with salt tolerance, including fresh and dry weights, plant height, relative water content, stomatal conductance. They were located on all chromosomes except chromosome 2 based on the alignment of their flanking sequences to the reference genome (*Medicago truncatula*). Of those identified, 13 were associated with multiple traits. Several loci identified in the present study were also identified in previous reports. BLAST search revealed that 19 putative candidate genes were collocated with 24 significant markers. Among them, B3 DNA-binding protein, Thiaminepyrophosphokinase and IQ calmodulin-binding motif protein were identified among multiple traits in the present and previous studies. With further investigation, the markers closely linked to drought and salt resistance can be used for MAS to accelerate the development of new alfalfa cultivars with improved resistance to drought and high salinity.

PO0763: Legumes, Soybean, Common Bean, and related

increase: Intelligent Collections of Food-Legume Genetic Resources for Agrofood Systems

Roberto Papa, Marche Polytechnic University (UNIVPM), Ancona, Italy

INCREASE is a European project (H2020-SFS-2019) focusing on the management and use of Genetic Resources on food legumes, which are crucial for sustainability, food security and human health. To meet this challenge INCREASE will expand the utilisation of food legumes genetic resources targeting users' needs in terms of accessibility, quality and quantity of information available. INCREASE, working with four important food legumes (chickpea, common bean, lentil, lupin) with significant value for the innovation of EU agriculture and food industry, will be based on four pillars: i) innovative data management solutions to develop gold standards for data sharing and integration into the central infrastructure, with decentralised data input, defined methodologies and best practices for exploitation of the novel information produced as well as the development of user friendly visualization tools; ii) developing novel tools and principles for germplasm management, based on the development of "Intelligent Collections" as a set of nested core collections of different sizes representing the entire diversity of each crop; iii) adoption of cutting-edge technologies for genotyping and phenotyping combined with the potential of Artificial Intelligence focusing on traits of interest for users; iv) international effort with the involvement of non-European partners and international organization to expand the scope and ambition of INCREASE. We will develop a citizen-science experiment, primarily aimed at dissemination of the project to stakeholders and citizens.

Overall, INCREASE will strengthen the field of legumes genetic resources and simultaneously it will represent an important model and tool for all crop genetic resources.

PE0764: Legumes, Soybean, Common Bean, and related

Efforts to Domesticate Hairy Vetch (*Vicia villosa* Roth): Eliminating Hard Seed and Pod Shatter

Lisa Kissing Kucek¹, Heathcliffe Riday¹, Allen N. Burke², Sarah S. Eagen³, Nancy Ehlke⁴, Sarah A. Krogman⁵, Steven B. Mirksy², Chris Reberg-Horton³, Matt R Ryan⁶, Michael Trammell⁵, Sandra Wayman⁶ and Nick P. Wiering⁴, (1)US Dairy Forage Research Center (USDA-ARS), Madison, WI, (2)USDA-ARS Beltsville Agricultural Research Center, (3)North Carolina State University, (4)University of Minnesota, (5)Noble Research Institute, LLC, (6)Cornell University

Hairy vetch, *Vicia villosa* (Roth), is a common cover crop that supplies nitrogen and improves water quality. Seed dormancy and pod dehiscence reduces seed yield and makes the species a weed. Thousands of years ago, domestication efforts eliminated hard seed and shatter in many agricultural crops. We report on a domestication effort to reduce hard seed and pod dehiscence in hairy vetch. Among seven locations and three years in the United States, we screened over 1600 genotypes for seed dormancy, pod dehiscence, and force required to cause dehiscence. Genotypes exhibited a complete range of variation in the traits of interest, from 0 to 100% seed dormancy and pod shatter. Despite left-skewed distributions of both traits, outlier genotypes were identified with soft seed and low dehiscence. As seed dormancy and pod shatter were controlled by a small number of genes in related species, heritability was expected to be high. However, heritability of seed dormancy and pod shatter was moderate in hairy vetch, at $h^2=0.3$ and 0.2 , respectively. Weather conditions during plant growth, maturity timing, post-harvest handling, and time between harvest and sample screening all influenced seed dormancy and pod shatter. Such influences likely reduced heritability, clouding the typically simple genetic mechanisms behind these traits. A correlation between seed yield and pod dehiscence ($r=0.0124$, $p=0.0052$) presented a tradeoff and additional challenge for selection.

PO0765: Maize, Sorghum, Millet, Sugar Cane, and related

Machine Learning Based Comparative Analysis of Gene Regulatory Networks in Monocots

Donat Wulf and Andrea Braeutigam, Bielefeld University, Bielefeld, Germany

Gene regulatory networks (GRN) control plant development in time and space and responses to environmental cues. The vast amount of publicly available RNAseq data can be harnessed with machine learning algorithms such as random forest prediction to assign transcription factors to their target genes. We hypothesized that evolutionary

comparisons in five monocot species will identify evolutionarily conserved transcription factors for core traits and changes in the GRN structure for derived traits.

The analysis is focused on photosynthesis as a core trait and C4 photosynthesis as a derived trait. The network of the species *Zea mays*, *Sorghum bicolor*, *Triticum aestivum*, *Hordeum vulgare* and *Oryza sativa* ssp. *Japonica* were inferred. We overcome the machine learning based high error rate in the prediction of targets by combining it with enrichment analyses. These enrichments determine transcription factors which are involved in the regulation of photosynthesis and C4 photosynthesis. By drawing comparisons between the five monocot species we are able to identify conserved regulatory patterns in these species. Comparing the C4 species *Zea mays* and *Sorghum bicolor* with the C3 species *Hordeum vulgare*, *Triticum aestivum* and *Oryza sativa* ssp. *Japonica* reveals that the C4 genes share the regulators with the conserved photosynthesis regulators in all species. Motifs were predicted for the putative C4 regulators based on network data.

PE0766: Maize, Sorghum, Millet, Sugar Cane, and related

Engineering Nitrogen Storage Capacity of Plant to Improve Drought Tolerance

Rajeev Gupta, ICRISAT, Patancheru, India and Kanwarpal S. Dhugga, INTERNATIONAL CENTER FOR MAIZE AND WHEAT IMPROVEMENT, EL BATAN, EM, Mexico

The global demand for nitrogen (N) fertilizer for agricultural production, which already stands at ~90 million metric tons per year, is projected to increase to 250 million metric tons by the year 2050. Nitrogen (N), the most abundant and expensive inorganic input for crop production, is mostly absorbed from the soil in the form of nitrate. When N is supplied in reduced form, soil microbes oxidize it to nitrate, which is delivered to the root surface by mass flow, that is, with the flow of water in the soil. Because it moves with water in the soil, nitrate is readily lost through leaching and run-off, contaminating underground water and causing excessive algal growth in fresh waterways and river deltas where dead zones develop, asphyxiating aquatic life. We posited that under normal growth conditions there might be periods in plant growth when the leaves produced excess photosynthate than could be utilized for optimal plant growth. The excess photosynthate could be sequestered into an osmoneutral form of reduced N, for example, vegetative storage proteins (VSPs), which could then be remobilized during recovery from transient stresses. VSPs have been known to accumulate in dicot plants when reproductive sink becomes limiting. In deciduous trees, N is remobilized from the leaves before they are shed in autumn and stored as VSPs in the phloem of the branches and the trunk. VSPs have not yet been reported in any of the grass species. We have been exploring various approaches in our efforts to improve stress tolerance. One of these is the addition of a transient N storage mechanism into the plant that would increase its N storage capacity, which could be useful for plant growth during periods of stresses. By subjecting maize seedlings to high N in the growth medium, we identified a lipoxygenase (Lox6) protein that exhibited the characteristics of a VSP. Lox6 was localized to chloroplasts which was confirmed by GFP fused to putative targeting signal peptide. Upon over-expression of *Lox6* under the control of various promoters (UBI, SSU, and PEPC) and targeting signals, an order of magnitude more Lox6 protein accumulated in the leaf tissue without any detectable, detrimental effect on the other, major leaf proteins. Whether it was expressed in the mesophyll or bundle sheath cells, the transgenic protein accumulated in the chloroplasts. Lox6 remobilized from the leaf like the other proteins during senescence. The hybrids overexpressing Lox6 driven by PEPC promoter was tested at three geographic locations under managed water stress and irrigated conditions. Grain yield of three of the transgenic events exceeded that of the control hybrid by a statistically significant 10-20% margin, providing evidence for our hypothesis that additional reduced form of N in the leaves can buffer a plant against abiotic stresses. In summary, our study provides evidence for the occurrence of a VSP-like protein in maize as well as its contribution to tolerance against water stress when overexpressed in the mesophyll cells. Our results provide a new avenue to improve drought tolerance in plants.

PO0767: Maize, Sorghum, Millet, Sugar Cane, and related

Investigation on the Genetic Basis of Alternation of Nutrient Stress Resilience during the Domestication of Grass Crops

Guangchao Sun¹, Nishikant Wase², James C Schnable³ and Toshihiro Obata², (1)Quantitative Life Sciences Initiative, University of Nebraska-Lincoln, Lincoln, NE, (2)Center for Plant Science Innovation,

University of Nebraska-Lincoln, Lincoln, NE, (3)Center for Plant Science Innovation, University of Nebraska-Lincoln, Lincoln, NE

Over the long term of domestication, grass crops breeding was biased towards high yield potential with optimized land management while potentially losing a number of resilient traits such as nutrient stress tolerance. Advancing our understandings on ancestor status of stress resilience in un-domesticated crop relatives will provide a delicate genetic reservoir for future crop breeding. In this study, we applied nitrogen and phosphorus starvation treatment with two highly domesticated crop maize and sorghum and their wild grass relative *Paspalum* and carried out comprehensive comparison with multi-omics data generated from the root tissue collected from the treated and untreated plants. Phenotypically, excessive root elongation and branching were consistently observed across three species under nitrogen starvation while no significant difference of these root traits by phosphorus depletion. The above ground part present obvious anthocyanin accumulation in stalks of maize and sorghum but not in *Paspalum*. In terms of biomass and primary metabolites, *Paspalum* showed the best resilience to both phosphorus and nitrogen starvation suggesting that resilience is an ancestor status of grass crops but lost over long time of domestication. Transcriptome analysis revealed that core nutrient stress responsive genes are predominantly involved in transporter activities and oxidoreductase activities likely to increase nutrient acquisition efficiency and antioxidation. Furthermore, most of the differentially regulated genes by moderate nutrient stress were syntenically conserved while severe nitrogen starvation would induce more differentially expressed nonsyntenic genes which are relative young suggesting ongoing development of species specific stress responsive networks. There is a set of syntenic genes of which the transcriptional response is highly conserved across all three species undertaking highest selection pressure in *Paspalum*. This suggests that domestication plays a role in relaxing nutrient stress selection on the core nutrient starvation responsive mechanisms. Together, along with a new genome assembly, we show that *Paspalum* as a wild relatives (un-domesticated) to maize and sorghum, is more resilient to nutrient stress which implies that domestication of grass crops resulted in loss of resilience to nutrient stress.

PE0768: Maize, Sorghum, Millet, Sugar Cane, and related

Pedigreed Mutant Library As an Efficient Resource to Discover Targets for Genome Editing

Zhanguo Xin¹, Gloria Burow², Chad Hayes², Yves Emendack², Junping Chen² and Doreen Ware³, (1)USDA ARS, Lubbock, TX, (2)USDA-ARS, Lubbock, TX, (3)Cold Spring Harbor Laboratory, Cold Spring Harbor, NY

Mutagenesis has been a classical and effective approach to dissect metabolic, developmental, and signal transduction pathways in both plants and animals due to the high density and random nature of induced mutations. A manageable number of lines can yield large number of mutations covering the entire genome. However, it suffers from two main disadvantages: 1) the lengthy and expensive process to identify the causal mutation and 2) the high density of background mutations. These two disadvantages have been largely overcome by the development in Next-Generation-Sequencing (NGS) and precise and efficient genome-editing technologies. We have established Pedigreed Mutant Library in the sorghum inbred line BTx623 by mutagenizing the seeds with ethyl methane sulfonate (EMS). This library has 6,400 M₄ seed pools and possesses a great diversity of mutant phenotypes. A large number of sorghum mutants with altered agronomic traits has been isolated from the mutant library. We have established an effective bioinformatic pipeline to identify the causal mutations through bulk-segregant-analysis (BSA) of the whole genome sequencing data of the pooled mutants selected from F₂ populations. Once an F₂ backcrossed population is established, the cost to identify the causal mutation is under \$300 in NGS sequencing. In the last few years, we have identified over 30 causal mutations underlying the altered agronomic traits. These causal mutations can serve as targets for genome-editing in elite sorghum lines. The combination of high throughput gene discovery from mutant library with precise genome editing will truly revolutionize plant and animal breeding.

PO0769: Maize, Sorghum, Millet, Sugar Cane, and related

Hosting a Plethora of Maize Genomes at MaizeGDB

Carson Andorf¹, Ethalinda Cannon¹, Jack M. Gardiner², Lisa Harper¹, Nancy Manchanda³, John Portwood¹, Mary Schaeffer⁴ and Margaret Woodhouse¹, (1)USDA-ARS, Ames, IA, (2)Division of

Animal Sciences, Columbia, MO, (3)Iowa State University, Ames, IA, (4)USDA ARS and University of Missouri, Columbia, MO

MaizeGDB currently holds hosts 18 genome assembly versions for 12 different maize accessions. At least 30 more are expected in the next year, including the 25 NAM parents and version 5 of the representative maize genome, B73. While these genome assemblies have great potential to advance maize research, hosting so many genomes in a manner that is beneficial for maize researchers and breeders is not straight-forward. Many assemblies also present the option for constructing pan-genomes and identifying pan-gene sets, as well as presenting challenges for handling gene model annotations from many different assemblies. Here we describe our approaches to handling many genome assemblies and our plans for a near future when the genome assemblies are likely to number in the hundred.

PE0770: Maize, Sorghum, Millet, Sugar Cane, and related

Genomic Data Mining with MaizeMine

Jack M. Gardiner¹, Md Shamimuzzaman¹, Justin J. Le Tourneau¹, Deborah A. Triant¹, Amy T. Walsh¹, Carson M. Andorf² and Christine G. Elsik¹, (1)University of Missouri, Columbia, MO, (2)USDA-ARS Corn Insects and Crop Genetics Research Unit, Iowa State University, Ames, IA

The availability of high-throughput genomic technologies has accelerated the generation of massive quantities of genomic datasets. Maize researchers often wish to perform comparative analysis between their own datasets and published or publicly available data. [MaizeMine](https://www.maizegdb.org/), the data mining warehouse for MaizeGDB (<https://www.maizegdb.org/>), enables researchers without scripting skills to integrate their data with publicly available data and perform meta-analysis. The MaizeMine List tool allows users to upload identifiers to create custom lists, perform set of operations such as unions and intersections, and execute template queries with lists. Users can easily compare their results with published results by uploading genomic coordinates or identifiers. MaizeMine uses the InterMine data warehousing system to integrate genomic sequences from the B73_RefGen_v3 and B73_RefGen_v4 genome assemblies, three sets of gene annotations (AGPv3, AGPv4, RefSeq), Gene Ontology (GO), protein annotations (UniProt), protein families and domains (InterPro), homologs (Ensembl Compara), and pathways (CornCyc, KEGG, Plant Reactome). In our most recent update of MaizeMine (v1.3), we have added data sets for a SNP array (Illumina SNP50), whole genome EMS mutagenesis sites, three insertional mutagenesis collections (Brutnell AcDs, Barker_Mu Illumina, and McCarty Uniform Mu), MaizeGamer annotations, and root and shoot transcriptional start sites. MaizeMine also provides pre-computed variant effects and expression levels based on RNA-seq data from the Zea mays Gene Expression Atlas (NCBI BioProject PRJNA171684). Database cross references facilitate easy gene identifier conversion between AGPv3, AGPv4 and RefSeq. MaizeMine provides simple and sophisticated search tools, including a keyword search, built-in template queries with intuitive search menus, and a QueryBuilder tool for creating custom queries.

PO0771: Maize, Sorghum, Millet, Sugar Cane, and related

Curation of Corn RefSeq to Provide a Robust Resource for Research and Annotation

Brian Smith-White, National Center for Biotechnology Information (NCBI/NLM/NIH), Bethesda, MD and Terence D. Murphy, NCBI/NLM/NIH, Bethesda, MD

The Reference Sequences (RefSeq) project at NCBI aims to provide a comprehensive set of well-annotated sequence records for a diverse set of organisms. RefSeq records serve as the foundation for NCBI's Gene resource, which provides an integrated view of gene information including gene transcript and protein products, genome annotation, publications demonstrating function of the gene product(s), information about expression, and links to pertinent resources both within NCBI and outside NCBI. Together, RefSeq and Gene provide information for nearly one-thousand eukaryote genomes, including a diverse set of agriculturally-important species. These data are available through a variety of NCBI resources including BLAST, and via FTP and NCBI's e-utilities APIs.

Corn has been targeted for additional in-depth curation of sequence and gene records by the NCBI staff and in collaboration with MaizeGDB. NCBI's curation efforts target a variety of potential gene annotation problems such as i) loci that are likely to be pseudogenes (both transcribed and untranscribed), ii) cultivar-specific sequences including genes not found in the B73 reference genome, and iii) incomplete or otherwise poorly annotated gene models. Efforts are made to assign informative gene and protein names in collaboration with MaizeGDB to

minimize nomenclature conflicts. To date, nearly 15,350 corn RefSeq Gene records have been curated, supplemented by a comprehensive gene annotation set for both protein-coding and non-coding loci produced by NCBI's Eukaryotic Gene Annotation Pipeline.

PE0772: Maize, Sorghum, Millet, Sugar Cane, and related

Sequence, Assembly and Annotation of Bayer's Maize Inbred Line LH244; A New Resource for Maize Genetics and Transformation

Zane A. Goodwin, Bayer Crop Science, Saint Louis, MO

- Bayer Crop Science, in collaboration with NRGene and the University of Wisconsin, reports the release of the LH244 inbred maize transformation line germplasm and assembled reference genome to academic research communities.
- The germplasm will be released to public seed stock centers and the assembled, annotated genome and a physiological description of the line will be published, and resources for plant transformation will be available at the University of Wisconsin Crop Innovation Center.
- LH244 is a commercially relevant inbred line that is readily transformable, thus making it a complete resource for genomic and genetic exploration.
- Here, we share insights into the unique features of the LH244 genome, transformability and physiology that make it a foundational resource for the maize genetics community.

PO0773: Maize, Sorghum, Millet, Sugar Cane, and related

Rapid, Heat-Induced Transgenerational Reactivation of a Silenced Transposable Element in Maize

Wei Guo, Purdue University, West Lafayette, IN

Transposons make up a substantial portion of most plant genomes. Due to their mutagenic potential, most of them are silenced. Although we know a lot about the means by which transposable elements are silenced, little is known about how their silencing status is maintained. Once reactivated, what accounts for a somatically or a transgenerationally transmitted reactivation state? Here, we use a minimal *Mutator* line that includes a naturally occurring variant of the *MuDR* transposon that can heritably trigger epigenetic silencing of that transposon. *MuDR* carries two genes *mudrA* and *mudrB*. We demonstrated that *Mediator of Paramutation1 (MOP1)*, a putative RNA-dependent RNA polymerase-encoding gene, is required for the maintenance of *mudrA* silencing. However, silenced *mudrA* is only progressively reactivated after multiple generations in a *mop1* mutant background. In contrast, *mudrB* never becomes reactivated. We find that all DNA methylation is lost at *mudrA* in *mop1* mutants in the first generation. Despite this, *mudrA* remains transcriptionally silenced in this generation. Remarkably, we find that this reactivation can be dramatically accelerated in seedlings carrying a silenced *MuDR* element in a *mop1* mutant background after a brief exposure to high temperature. In contrast to previous observations, in heat stressed plants, both *mudrA* and *mudrB* are reactivated. This active state is maintained throughout the life of the plant after the initial trigger has disappeared. Remarkably, this activity is transmitted to the next generation and is not associated with DNA methylation. This is intriguing because *mudrA* and *mudrB*, which are associated with two mutually exclusive histone marks, are both reactivated upon heat exposure, suggesting that heat stress might integrate two distinct epigenetic pathways to wake up a silenced transposon. Taken together, this project will give us an opportunity to better understand the maintenance of transposon silencing, and the relationship between epigenetic silencing and stress response.

PE0774: Maize, Sorghum, Millet, Sugar Cane, and related

The Effects of Epigenetic Modifications and Active Transposable Elements on Meiotic Recombination in Maize

Meixia Zhao¹, Tyshawn Ferrell¹, Diya Yang¹, Claire Stoll¹, Chung-Ju Rachel Wang² and Damon Lisch³, (1)Miami University, Oxford, OH, (2)IPMB/Academia Sinica, Taipei, Taiwan, (3)Purdue University, West Lafayette, IN

Meiotic recombination generates genetic diversity and ensures the accurate segregation of homologous chromosomes. Meiotic crossovers are not uniformly distributed along chromosomes. There are hotspots in the hypomethylated chromosomal arms and suppression in heterochromatic regions. The causes of this variation is not well understood, but epigenetic modifications such as DNA methylation and chromatin states are thought to be important factors. In maize, we have examined the effects on meiotic recombination of mutations in a component of the RNA directed DNA methylation pathway, *MOP1* (*mediator of paramutation1*), a putative RNA-dependent RNA polymerase, and a key component in *trans*-acting small RNA pathway *LBL1* (*leafbladeless1*). We found that meiotic recombination was uniformly decreased in the heterochromatic regions of all the 10 maize chromosomes but was increased in most of the examined euchromatic regions in *mop1* mutants. Interestingly, we also found DNA methylation may affect sex-specific differences in the frequency of meiotic recombination, suggesting that these differences may have an epigenetic component. In contrast, no significant changes in the frequency of meiotic recombination in both the euchromatic and heterochromatic regions were observed between *lbl1* mutants and the wild type plants. Additionally, meiotic recombination is initiated by the formation of DNA double strand breaks (DSBs) and is completed by the repair of these breaks during meiosis. Given that TEs can be a source of DSBs as well, we hypothesize that widely scattered active *Mutator* (*Mu*) transposons can generate SPO11-1 independent DSBs and stimulate overall increases in recombination. We have developed several F2 and backcross populations to compare recombination with and without *Mu* TEs, with and without *Mu* activity, and with and without *SPO11* and *MOP1*.

PO0775: Maize, Sorghum, Millet, Sugar Cane, and related

Discovery of Structural Variation with Whole Genome Sequences and Bionano Genome Maps in Maize

Ha Le¹, Guifang Lin¹, Pierre A Migeon², Jie Ren³ and Sanzhen Liu¹, (1)Kansas State University, Manhattan, KS, (2)Kansas State University Department of Plant Pathology and Interdepartmental Genetics, Manhattan, KS, (3)Integrated Genomics Facility, Manhattan, KS

High levels of structural variation (SV) are known to exist between maize lines, and yet much of this variation has proven difficult to examine. The physical maps of high-molecular-weight DNA generated by Bionano genome mapping technology offer a new approach for SV discovery. We produced Bionano maps for two maize inbred lines B73 and Mo17. The comparison of these two maps, in combination with assembly comparison, read-depth analysis with whole genome sequencing data, and genetic mapping of SV with sequencing data of segregation individuals revealed a relatively complete set of SV between B73 and Mo17, including a previously discovered ~2.6 Mb deletion in Mo17. Sequencing data from a segregating population of B73xMo17 hybrids was used to support and map the SVs detected using the read-depth and Bionano datasets. We also used the updated Bionano genome mapping technology that yields a higher continuity of genome maps for another maize inbred line, A188. The maps have dramatically improved the A188 genome assembled with long Nanopore reads. The maps also resolved the *White Cap* (*Wc*) locus in A188, which contains high-copy tandem repeats that were not resolved with Nanopore reads. Each *Wc* repeated unit, approximately 27 kb in size, carries a copy of *ccd1* that encodes carotenoid cleavage dioxygenase. Higher copy number of *ccd1* is associated with reduced levels of carotenoid in the endosperm. In summary, emerging technologies such as Bionano genome mapping offer new tools to further reveal the complexity of maize genomes, which will be essential to fully characterize genomic variation among maize lines and their functional impacts.

PE0776: Maize, Sorghum, Millet, Sugar Cane, and related

Genetic Diversity and Population Structure in Sweet Maize

Lillian M. Hislop, University Of Wisconsin-Madison, Madison, WI, William F. Tracy, University of Wisconsin-Madison, Madison, WI, Michael A. Gore, Plant Breeding and Genetics Section, School of Integrative Plant Science, Cornell University, Ithaca, NY and Matt Baseggio, Seneca Foods Corporation, LeSueur, MN

Genetic diversity is vital to maintain the effectiveness of breeding programs at achieving goals to improve products and adapt to developing abiotic stresses. Germplasm collections contain the necessary phenotypic and genotypic

diversity; however, characterization of these resources is necessary to utilize them fully within breeding programs. The goal of this experiment was to identify key components of the population structure within the available sweet maize breeding germplasm. We used an Illumina NextSeq500 to genotype 67126 sites using Genotyping By Sequencing for 439 accessions. The accessions were selected as a representative subset from the major US sweet maize breeding programs. STRUCTURE and principal component analysis identified major subpopulations within the collection, differentiated by the endosperm mutant and the breeding program of origin. Phylogenetic analysis confirmed the accuracy of the established pedigree information. This research demonstrates the magnitude and sources of diversity found within available sweet maize germplasm, which will be helpful for informing sweet maize breeding decisions.

PO0777: Maize, Sorghum, Millet, Sugar Cane, and related

Influence of Selection on Performance Stability in the Maize BSSS Population

Bridget McFarland¹, Shawn Kaeppler¹ and Natalia de Leon², (1)University of Wisconsin-Madison, Madison, WI, (2)University of Wisconsin - Madison, Madison, WI

To ensure dependable plant growth and productivity in farmers' fields, plant breeders evaluate the stability, or consistency of plant performance across a range of environments, and then select stable hybrids that maximize productivity. The goals of this study were 1) identify environmental stability of hybrids generated by unselected, less selected, and more recently released inbreds derived from the Iowa Stiff Stalk Synthetic (BSSS) population and 2) evaluate how estimates of stability can be refined using climatic data. A set of hybrids was generated by crossing 102 inbred lines derived from BSSS (which varied in release date and selection level) by tester DK31IH6. Hybrids were evaluated in 31 environments as part of the GenomesToFields Initiative across two years. Stability was estimated using two environmental indices with Finlay-Wilkinson linear regressions; (1) performance-based, using the average hybrid performance in each environment; and (2) environmentally-based, using average hybrid photothermal time at flowering in each environment. Slope and mean squared error (MSE) estimates were extracted. Estimates from both indices suggest that selecting for improved plant performance is associated with improved stability. The average population differentiation index (F_{ST}) indicates that selecting for productivity has reduced genetic similarities of the founding BSSS inbred lines as represented on the more selected lines. When comparing values' significant difference from ideal stability, the environmentally-based index identified more hybrids as stable compared to the performance-based index. Future work will focus on identifying key environmental factors to optimize performance stability estimation and ultimately assist breeders with evaluations across environments.

PE0778: Maize, Sorghum, Millet, Sugar Cane, and related

Genome-Wide Analysis of F1 Hybrids and F2 Progeny to Determine the Initiation and Maintenance of Epigenetic Silencing in Maize

Diya Yang¹, Meixia Zhao¹, Dafang Wang² and Damon Lisch³, (1)Miami University, Oxford, OH, (2)Delta State University, Cleveland, MS, (3)Purdue University, West Lafayette, IN

The maize genome is largely composed of transposable elements and other repeats, most of which are silenced by epigenetic marks. We wished to determine the initiation and maintenance of epigenetic silencing of the repeats using two mutants involved in the small RNA biogenesis pathways. These include *mop1* (*mediator of paramutation1*), a putative RNA-dependent RNA polymerase in RNA-directed DNA methylation pathway, and *lhl1* (*leafbladeless1*), a key component in *trans*-acting small RNA pathway. To determine the initiation of silencing, we performed high-throughput whole genome bisulfite, small RNA and mRNA sequencing of the F1 hybrid plants of the mutant as well as of wild type siblings in two inbred maize lines B73 and Mo17. Consistent with previous results, our analysis of the small RNA sizes and abundance in *mop1* revealed that 24 nucleotide small RNAs were dramatically reduced in the *mop1* mutants compared with the wild type plants. "CHH islands" in the upstream and downstream regions flanking genes were completely removed, demonstrating a significant role of *mop1* in *de novo* CHH methylation. In contrast, no significant global changes of removal of DNA methylation were observed in the *lhl1* F1 mutants. Our next step will be to identify all the heterozygous differentially methylated regions in the mutants compared with wild type plants between these two lines, and to investigate whether the transfer of methylation can be disrupted by the mutants. To determine the maintenance of the epigenetic silencing triggered by one genetic background to another genetic background, we will investigate the heritance in the F2 progeny derived from the F1 hybrid mutants and the wild type plants.

PO0779: Maize, Sorghum, Millet, Sugar Cane, and related

A Step Forward in Predicting Phenotypic Performance in Diverse Field Environments

Qi Mu, Xianran Li, Tingting Guo, Xin Li and Jianming Yu, Iowa State University, Ames, IA

Phenotypic plasticity describes that a genotype behaves differently when exposed to different environments. Phenotypic plasticity creates difficulty for plant breeders to select the best breeding lines across locations and years. Understanding and being able to predict phenotypes across diverse environments will facilitate selecting stable or adaptive lines with fewer amount of resources. Recently, we established a joint genomic regression analysis (JGRA) framework to dissect the complex flowering time plasticity observed in natural field environments by leveraging an explicit environmental index. This is a critical step forward to implement real environmental indexes into phenotypic plasticity predictions. In this study, we hypothesized that plant height plasticity can be unraveled and explained in a similar manner. The objectives were to 1) uncover the patterns of sorghum plant height plasticity in diverse environments; 2) predict performance in new environments; 3) identify and dissect the genetic determinants to explain the observed plasticity. Our results showed that varied degree of plasticity in plant height of sorghum lines could be explained, modeled, and predicted with a biologically meaningful environmental index. High prediction accuracy was achieved by using this environmental index. The effects of three height QTLs changed dynamically across environments, contributing to the observed phenotypic plasticity. In conclusion, integrating environmental data with genomic components has enhanced our understanding of phenotypic plasticity, and enabled predictive modeling for multiple agronomically important traits under diverse field environments.

PE0780: Maize, Sorghum, Millet, Sugar Cane, and related

Assessing the Impact of Different Genetic Models to Determine Heterotic Patterns in Tropical Maize

Julia Silva Morosini¹, Fernando Garcia Espolador¹, Miriam Suzane Vidotti², Filipe Couto Alves³, Raysa Gevartosky⁴, Anna Rita Marcondes dos Santos⁴, Natalia de Leon¹ and Roberto Fritsche Neto⁴, (1)University of Wisconsin - Madison, Madison, WI, (2)Plant Breeding Department, Federal University of Goiás, Goiás, Brazil, (3)Department of Epidemiology and Biostatistics, Michigan State University, East Lansing, MI, (4)Genetics Department, “Luiz de Queiroz” College of Agriculture, University of São Paulo, Piracicaba, Brazil

Commercial maize breeding relies mainly on heterosis, a phenomenon best explored when heterotic groups are assigned. Although the use of molecular marker-based relationship matrices tends to improve the predictability of genetic effects, dominance is often neglected. In this study, we considered models including additive and the combination of additive+dominance effects for the estimation of combining abilities and the determination of heterotic groups in tropical maize. For that, 906 single-crosses obtained from a diallel scheme of 49 inbred maize lines were genotyped *in silico* using 34,571 SNP and evaluated for grain yield in four environments in São Paulo, Brazil. Three modeling scenarios were considered: (A) pedigree-based, (B) additive, and (C) additive+dominance effects. We observed an increment of additive variance when markers were used to calculate a relationship matrix. There was an increasing efficiency in capitalizing dominance variance throughout the models. The ranking of lines was highly similar among scenarios, which indicates that pedigree alone deduced within the diallel analysis was sufficient to capture allelic effects. The specific combining abilities assigned distinct clustering of parents for each scenario. More considerable differences in pool composition occurred between scenarios B and C, although there was still a meaningful overlap. Our findings reveal the additive effect is well assessed by using pedigree data when the population lacks structure. In addition, additive kernel accounted for most of the genetic variance, which corroborates the critical role of the additive-by-additive epistasis in heterosis expression. Finally, the incorporation of dominance provides relevant information for the determination of heterotic groups.

Funding acknowledgment: São Paulo Research Foundation (FAPESP) - grant#2017/25549-6 and grant#2019/06532-0; Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001; University of Wisconsin-Madison.

PO0781: Maize, Sorghum, Millet, Sugar Cane, and related

Development and Genetic Characterization of a Tropical Maize Diversity Panel

Fernando Garcia Espolador¹, Julia Silva Morosini¹, Humberto Fanelli Carvalho², Dayane Cristina Lima¹, José Felipe Gonzaga Sabadin³, Natalia de Leon¹ and Roberto Fritsche Neto³, (1)University of Wisconsin - Madison, Madison, WI, (2)Agronomic Institute of Campinas, Campinas, Brazil, (3)Genetics Department, “Luiz de Queiroz” College of Agriculture, University of São Paulo, Piracicaba, Brazil

Tropical maize genotypes exhibit high genetic diversity and are an important source of potential alleles for breeding programs. Despite their usefulness as an active genetic resource, their utilization is limited due to the lack of widespread genetic characterization of such collections. In this context, this study seeks to assemble and genetically characterize a tropical maize diversity panel. For that, 360 highly diverse inbred lines from two distinct active germplasm banks were utilized. They were genotyped using a genotyping-by-sequencing approach with two restriction enzymes (PstI and MseI) and aligned to the version 4 of the B73 reference genome. For genetic assessments, two datasets were considered: one with the raw data, and the second one with data imputed and filtered for quality control, retaining only biallelic markers with minor allelic frequency > 0.05, call rate > 0.95, and linkage disequilibrium (r^2) < 0.99. The raw dataset contained 167,194 SNPs well distributed across the chromosomes. The G/C and A/T ratios were close to 1. The proportion of missing data was 0.36, and the average observed heterozygosity was 0.03. The filtered dataset, containing 14,560 SNPs, showed similar estimates of populational genetic parameters compared to the first one. A clustering analysis suggested that this panel contains well-structured groups and confirmed the high genetic diversity present in it. Our initial results support the utility of this maize diversity panel to facilitate a deeper understanding of tropical germplasm. The information derived from this study can potentially assist in establishing future breeding programs and scientific discoveries.

Keywords: Germplasm, Corn, structure, diversity, Brazil

Funding acknowledgment: This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001, National Council for Scientific and Technological Development (CNPq), São Paulo Research Foundation – FAPESP (Grant 2017/24327-0), and University of Wisconsin-Madison.

PE0782: Maize, Sorghum, Millet, Sugar Cane, and related

Genetic and Genomic Resources for Maize Transformation

Guifang Lin¹, Yan Liu², Tej Man Tamang¹, Jun Zheng², Mingxia Zhao¹, Ha Le¹, Cheng He¹, Junjie Fu², Myeong-Je Cho³, Sunghun Park¹, Hairong Wei⁴, Frank F. White⁵, Yunjun Liu², Guoying Wang² and Sanzhen Liu¹, (1)Kansas State University, Manhattan, KS, (2)Chinese Academy of Agricultural Sciences, Beijing, China, (3)Innovative Genomics Institute, Sunnyvale, CA, (4)Michigan Technological University, Houghton, MI, (5)University of Florida, Gainesville, FL

Genome editing tools provide great potential for the elucidation of gene functions and crop improvements. In maize, however, current genetic and genomic resources as well as transformation capacity are insufficient for the full utilization of editing tools. We generated Nanopore long sequencing reads and BioNano physical mapping data to produce a reference-level genome assembly of a highly transformable maize inbred line A188. Genetic mapping using an F2 population of A188 and B73, an elite inbred line recalcitrant to callus culture or transformation, identified at least five genomic loci associated with callus culturability. In addition, hundreds of double haploids (DHs) and recombinant inbred lines (RILs) from the cross of A188 x B73 were generated, which will be used for validation of five association loci and selection for highly amenable maize lines with different combinations of the two genomes. Furthermore, using RNA sequencing data from diverse tissues of A188, including callus samples from multiple stages, we will construct gene regulatory networks to identify highly hierarchical regulators and hub genes that govern maize embryogenesis.

P00783: Maize, Sorghum, Millet, Sugar Cane, and related

Genome-Wide Association Study of Maize Flowering Time with K-Mer Read Counts

Cheng He, Yangfan Hao and Sanzhen Liu, Kansas State University, Manhattan, KS

Genome wide association study (GWAS) with single nucleotide polymorphisms (SNPs) has been widely used to explore genetic architectures of quantitative traits. The reliance on alignments of sequencing reads against reference genomes introduces bias and limits examination only on accessible low repetitive genomic regions. Here we employ a k-mer based approach, an alignment-independent method, for genetic association studies. Specifically, counts of k-mers were determined from whole genome shotgun data of 269 maize inbred lines and associated with the trait of flowering time. With this method, we identified 863,648 flowering time associated k-mers. Co-occurrence network analysis using the top10k associated k-mers identified k-mer clusters with strong connectivity. Some clusters contained k-mers that were largely mapped to chr2, chr8, and chr10. Comparing with previous flowering time GWAS study based on SNPs, few of these k-mers were located in known flowering time related genes. Using B73 RNA-Seq of 739 samples from different tissues and development stages, we found 3,412 flowering associated k-mers co-expressed with genes in the photosynthesis pathways, leading to find two candidate genes for further functional validation. In summary, k-mer counting based GWAS provided an alternative strategy for association mapping and may uncover genes missed in traditional approaches.

PE0784: Maize, Sorghum, Millet, Sugar Cane, and related

Identification of Genetic Determinants of Maize Tassel Structure Using RGB Images

Yan Zhou¹, Seyed Vahid Mirnezami², Lakshmi Attigala¹, Srikant Srinivasan³, Baskar Ganapathysubramanian² and Patrick S. Schnable⁴, (1)Iowa State University, Ames, IA, (2)Department of Mechanical Engineering, Iowa State University, Ames, IA, (3)IIT Mandi, Mandi, India, (4)Department of Agronomy, Iowa State University, Ames, IA

Structural variation in inflorescence traits of cereal crops can influence yield. To identify genetic factors that contribute to the structural variation of maize tassels, we phenotyped tassels from a diversity panel consisting of 339 inbreds using an platform that generated RGB images of each tassel from five different camera orientations. An automated feature extraction pipeline was developed to measure five tassel traits from the resulting images. Comparisons with manually collected ground truth established that phenotyping accuracy varies among camera orientations in a trait-dependent manner. A novel binary trait, “tassel openness” was defined as whether or not the central spike is occluded by tassel branches. A genome-wide association study (GWAS) conducted for this trait identified multiple trait-associated-SNPs (TAS); two of which are adjacent to known inflorescence genes.

PO0785: Maize, Sorghum, Millet, Sugar Cane, and related

Using Multi-Omics to Dissect Maize Heterosis

Peng Zhou¹, Zhouxin Shen², Nathan M. Springer¹ and Steven Briggs², (1)University of Minnesota, St. Paul, MN, (2)University of California San Diego, La Jolla, CA

We compared proteomics data from 23 tissues in maize inbred lines B73, Mo17 and the hybrid with published transcriptomics data in order to elucidate heterosis at the molecular level

PE0786: Maize, Sorghum, Millet, Sugar Cane, and related

DNA Methyloome Selection Reshapes Gene Regulation and Affects Adaptation during Maize Domestication

Gen Xu¹, Jing Lyu¹, Qing Li^{2,3}, John F. Doebley⁴, Dafang Wang⁵, Nathan M. Springer⁶, Jeffrey Ross-Ibarra⁷ and Jinliang Yang¹, (1)University of Nebraska-Lincoln, Lincoln, NE, (2)HuaZhong Agriculture University, Wuhan, China, (3)University of Minnesota, Saint Paul, MN, (4)University of Wisconsin-Madison, Madison, WI, (5)Delta State University, Cleveland, MS, (6)University of Minnesota, St. Paul, MN, (7)University of California, Davis, CA

DNA methylation, a ubiquitous feature of plant genomes, plays a critical role in controlling gene expression and plant development, and likely affecting phenotypic variations. However, the landscape and variation of DNA methylation during maize domestication and improvement processes remain largely unknown. Here we leveraged whole-genome sequencing (WGS) and whole-genome bisulfite sequencing (WGBS) data on a natural population of teosinte (*Zea mays* ssp. *parviglumis*) and a set of modern maize lines and Mexican landraces to address population

epigenetic questions. We identified population-wide differentially methylated regions (DMRs) and found CG and CHG DMRs were largely enriched in the low recombination regions of the genome. Further investigation revealed that CG DMRs were enriched in the 5' untranslated regions and genes exhibiting CG DMRs were likely exhibiting binding activities. For several trait-associated DMRs, functional analyses suggested that these DMRs likely serve as cis-acting loci in regulating gene expression. Our results enabled a better understanding of the evolutionary forces acting on patterns of DNA methylation during maize domestication.

PO0787: Maize, Sorghum, Millet, Sugar Cane, and related

Characterizing and Fine-Mapping a Mutation in Maize which confers Reduced Leaf Area

Marlee Rose Labroo, University of Illinois at Urbana-Champaign, Urbana, IL

Improved varietal tolerance to high planting densities has been a major driver of historical maize yield increases in the United States, and planting density tolerance is strongly correlated with reduced leaf area. A recessive mutation which confers moderately reduced leaf area, *rdla*, may be useful in targeted improvement of yield at high planting densities and narrow row spacing in maize. Characterization of the mutation in an Oh43 inbred background suggests that the reduced leaf area phenotype is only apparent in adult leaves, and adult leaf area of *rdla* genotypes is reduced by approximately one-third compared to wild-type. The mutant plants were also assessed for differences in leaf number, height, stem diameter, root volume, specific leaf area, and cell size. The genomic interval containing the *rdla* mutation was narrowed to an ~4 Mb region on chromosome 4 through a mapping-by-sequencing approach with imputation in BC₅S₁₀ and BC₆S₁ individuals. The location was further narrowed to a ~70 Kb region of nine genes using an expanded BC₆S₁ fine-mapping population. Differential expression analysis and variant effect prediction to determine the causative *rdla* mutation are in progress. Marker-assisted backcrossing of *rdla* into expired plant variety protection lines was also initiated. These elite genotypes will then be used to assess the penetrance of the *rdla* mutation in both inbred and hybrid backgrounds, as well as any changes in planting density tolerance conferred by the mutation.

PE0788: Maize, Sorghum, Millet, Sugar Cane, and related

Effects of CENH3 Modifications on Its Deposition in Maize

Jing Yuan, State Key Laboratory of Plant Cell and Chromosome Engineering, Institution of Genetics and Developmental Biology - Chinese Academy of Sciences, Beijing, China

The centromere, as an essential element to mediate chromosome segregation, is epigenetically determined by CENH3-containing nucleosomes as a functional marker; therefore the accurate deposition of CENH3 is crucial to chromosome transmission. We characterized the deposition of CENH3 in maize by over-expression and mutational analysis. Our results revealed that over-expressing CENH3 in callus is lethal while over-expressing GFP-CENH3 and CENH3-YFP in callus and plants is not and can be partly deposited normally. Different mutations of GFP-CENH3 demonstrated that CENH3-Thr4 in the N terminus was needed for the deposition as a positive phosphorylation site and the last five amino acids in the C terminus are necessary for deposition. The C terminal tail of CENH3 is confirmed to be responsible for the interaction of CENH3 and histone H4, which indicates that CENH3 maintains deposition in centromeres via interacting with H4 to form stable nucleosomes. For GFP-CENH3 and CENH3-YFP, the fused tags at the termini probably affect the structure of CENH3 and reduce its interaction with other proteins, which in turn could decrease proper deposition. Taken together, multiple amino acids or motifs were shown to play essential roles in CENH3 deposition, which is suggested to be affected by numerous factors in maize.

PO0789: Maize, Sorghum, Millet, Sugar Cane, and related

Genetic Architecture of Composition Important for Corn Chip Production

Michael J. Burns, Amanda Gilbert, Jonathan Renk and Candice N. Hirsch, Department of Agronomy and Plant Genetics, University of Minnesota, St. Paul, MN

Nixtamalization is an alkaline cooking process used in the production of corn chips and tortilla products. The cooked kernels, referred to as nixtamal, need to have uniform moisture profiles within batches to ensure proper processing and quality control of the final product. There are traits that are hypothesized to have a large effect on

moisture uptake during nixtamalization, and understanding the genetic architecture of these traits is valuable to future breeding efforts targeted at producing superior food grade corn hybrids. We are using a mutagenesis approach to identify genes contributing to compositional variation for total protein, starch, fiber, oil, and sugars. To conduct the mutagenesis, ethyl methanesulfonate (EMS) was applied to the pollen of two inbred lines, PH207 and Oh43, followed by self pollination with the mutagenized pollen. Two more generations of self pollination were conducted to produce segregating ears within mutant families. Kernels from an entire ear were ground and NIR scanned to predict compositional attributes on a per ear basis for each of the compositional traits based on previously established NIR equations. For each family, 10 ears were evaluated. Potential segregating families for each trait were identified by the coefficient of variation among the individuals in the family. These segregating families will be validated in Summer 2020 followed by bulk segregant analysis of pooled DNA from the mutant and non-mutant individuals. Having a better understanding of the genes contributing to phenotypic variation for ear compositional attributes will allow breeders to use molecular breeding techniques to improve their populations.

PE0790: Maize, Sorghum, Millet, Sugar Cane, and related

Investigating Co-Expression Networks and Gene Regulation in Corn Roots Exposed to Weeds

David P. Horvath, USDA-ARS, Fargo, ND, Jigar Desai, North Carolina State University, Raleigh, NC and Colleen Doherty, NCSU, Raleigh, NC

Weeds reduce corn yield by altering corn developmental processes. In related studies, both below-ground and above ground signals were implicated in initiating these developmental responses that result in yield loss. To investigate early signaling processes involved specifically in below ground signals, corn plants were grown in pots and surrounded by a competitor (winter canola) that were planted in cone-tainers such that no root-to-root contact was possible. Likewise, an opaque barrier was placed between the canola and the corn so that light quality signals were blocked. When corn was 4 weeks old (V6 stage of development), the canola plants were removed from the cone-tainers and replanted in the resulting hole in direct root-to-root contact. Root tissue from mock-treated (canola was placed back into the cone-tainers and placed back into the soil) or treated corn plants were collected at 0, 1, 2, 3, 7, and 14 days and subjected to RNAseq analysis with three technical replicates for each treated and mock-treated time points, and this experiment was repeated twice. Differential gene expression and network analysis using WCGNA as well as a method to identify regulatory elements were identified from each experimental run independently and the overlapping set was characterized. We report here the results of these experiments.

PO0791: Maize, Sorghum, Millet, Sugar Cane, and related

Genetic Diversity and Population Structure Analysis of *Saccharum* and *Erianthus* Genera Using Microsatellite (SSR) Markers

Yong-Bao Pan¹, Ahmad Ali², Qinnan Wang³, Jinda Wang², Jun-Lv Chen³ and Sanji Gao², (1)USDA-ARS, SRU, Houma, LA, (2)Fujian Agriculture and Forestry University, Fuzhou, China, (3)Guangzhou Sugarcane Industry Research Institute, Hainan Sugarcane Breeding Station, Sanya, China

In order to understand the genetic structure and diversity within and among wild species of the highly complex polyploid genus of *Saccharum* and *Erianthus*, 79 accessions from five *Saccharum* species (*S. officinarum*, *S. spontaneum*, *S. robustum*, *S. barberi*, and *S. sinense*), 30 *Saccharum* spp. hybrids, and six accessions of *E. arundinaceus* were analyzed using 21 pairs of fluorescence labeled highly polymorphic SSR markers. A total of 167 polymorphic SSR alleles were identified using capillary electrophoresis (CE) detection system with a mean polymorphic information content (PIC) value of 0.92. Genetic diversity analyses involving number of polymorphic loci (NPL), percentage of polymorphic loci (PPL), number of observed alleles (*N_a*), number of effective alleles (*N_e*), Shannon's index (*I*), and Nei's gene diversity (*h*) revealed that *Saccharum* spp. hybrids were more diverse than *Saccharum* and *Erianthus* species. Based on the SSR data, the 115 accessions were classified into seven groups, which corresponded to the *Saccharum* and *Erianthus* accessions through phylogenetic analysis and principle component analysis (PCA). We propose that seven core SSR primer pairs, namely, SMC31CUQ, SMC336BS, SMC597CS, SMC703BS, SMC24DUQ, mSSCIR3, and mSSCIR43, may have a wide applicability in genotype identification of *Saccharum* species including *Saccharum* spp. hybrids. Thus, the information from this study contributes to the management of sugarcane genetic resources.

PE0792: Maize, Sorghum, Millet, Sugar Cane, and related

Characterization of the Sugarcane Genome By Oligo-FISH

Kai Wang, Fujian Agriculture and Forestry University, Fuzhou, China

Sugarcane (*Saccharum* spp., Poaceae) is a leading crop for sugar production providing 80% of the world's sugar. However, the genetic and genomic complexities of this crop such as its high polyploidy level and highly variable chromosome numbers have significantly hindered the studies in deciphering the genomic structure and evolution of sugarcane. Here, we developed the first set of oligonucleotide (oligo)-based probes based on the *S. spontaneum* genome ($x = 8$), which can be used to simultaneously distinguish each of the 64 chromosomes of octaploid *S. spontaneum* SES208 ($2n = 8x = 64$) through fluorescence in situ hybridization (FISH). By comparative FISH assay, we confirmed the chromosomal rearrangements of *S. spontaneum* ($x = 8$) and *S. officinarum* ($2n = 8x = 80$), the main contributors of modern sugarcane cultivars. In addition, we examined a *S. spontaneum* accession, Np-X, with $2n = 40$ chromosomes, and we found that it was a tetraploid with the unusual basic chromosome number of $x = 10$. Assays at the cytological and DNA levels demonstrated its close relationship with *S. spontaneum* with basic chromosome number $x = 8$ (the most common accessions in *S. spontaneum*), confirming its *S. spontaneum* identity. Population genetic structure and phylogenetic relationship analyses between Np-X and 64 *S. spontaneum* accessions revealed that Np-X belongs to the ancient Pan-Malaysia group, indicating a close relationship to *S. spontaneum* with basic chromosome number of $x = 8$. This finding of a tetraploid *S. spontaneum* with basic chromosome number of $x = 10$ suggested a parallel evolution path of genomes and polyploid series in *S. spontaneum* with different basic chromosome numbers.

PO0793: Maize, Sorghum, Millet, Sugar Cane, and related

Insights into Smut Whip Development in Sugarcane

Deepak Sehgal¹, Hugo V. S. Rody² and Claudia B. Monteiro-Vitorello¹, (1)University of São Paulo, Piracicaba, Brazil, (2)University of São Paulo, Piracicaba, Sao Paulo, Brazil

Sporisorium scitamineum, the causal agent of sugarcane smut disease, is a biotrophic filamentous fungi that brings out drastic changes in shoot architecture of sugarcane and results in mild to severe yield loss. Unraveling the key components and their interacting partners of this plant-pathogen interaction will help in better illustration of this complex interactive network leading to whip development in sugarcane. In a recent study, we observed some of the key floral integrator genes like AP1, Col6, and FT differentially expressed in a comparative transcriptional profiling of smut infected and mock plants, which navigated our research work to focus on the transcriptome profile of flowering pathway genes in smut infected and flowering plants. We identified 477 flowering gene orthologs in sugarcane with 68% similarity at the protein level with Arabidopsis and Rice. An *in silico* analysis of the interacting partners of some of the differentially expressed genes like LATE, TOE3, and AHL22 from autonomous, photoperiod and circadian clock pathways suggested how smut fungus initially activates the reproductive pathway after infecting the plant and later on mischievously employs a complex network of transcription factors/regulators to gain control of flowering pathway resources to develop whip.

PE0794: Maize, Sorghum, Millet, Sugar Cane, and related

Molecular Dissection of Agronomic Traits in *Saccharum* Spp. Hybrids

Md Sariful Islam¹, Xiping Yang², Sushma Sood³, Jack Comstock⁴, Fenggang Zan⁵ and Jianping Wang², (1)Sugarcane Production Research Unit, USDA-ARS, Canal Point, FL, (2)Agronomy Department, University of Florida, Gainesville, FL, (3)Sugarcane Production Research Unit, USDA ARS, Canal Point, FL, (4)Sugarcane Field Station, USDA-ARS, Canal Point, FL, (5)Yunnan Sugarcane Research Institute, Kaiyuan, China

Sugarcane is one of the most important crops in the tropical and sub-tropical regions worldwide. The main goal of sugarcane breeding programs is releasing new cultivars with improved sugar content, disease resistance and agronomic traits. Molecular markers linked to the sugar yield would greatly facilitate the development of sugarcane cultivars with higher sugar content. In this study, quantitative trait loci (QTL) associated with sugar and yield related traits were identified using a segregating F1 population derived from two *Saccharum* spp. hybrids. Specifically, BRIX, POL, recoverable sugar content (SC), fiber content (FC), moisture content (MC), juice purity, stalk diameter

(SD), and stalk weight (SW) data were collected from a replicated field trial of a bi-parental population. A total of 36 and nine QTL for sugar and yield related traits, respectively were identified using a high density genetic map with markers developed by genotyping-by-sequencing. Of the 45 detected QTL, seven QTL were associated with each of the three sugar related traits BRIX, POL, and SC; six QTL with FC and MC; three QTL with juice purity; four QTL with SD; and five QTL with SW. The QTL explained a total of phenotypic variations of 70.90, 61.80, 61.68, 68.67, 91.62, 33.00, 49.91, and 64.49 % for BRIX, POL, SC, FC, MC, purity, SD, and SW, respectively. Upon validation, markers from the identified QTL would be useful in marker-assisted selection for selecting superior cultivars with these traits.

PO0795: Maize, Sorghum, Millet, Sugar Cane, and related

Analysis of Genetic Diversity Among Sugarcane Genetic Resources in Thailand

Tatsuro Kimura, Agriculture & Biotechnology Business Div., TOYOTA MOTOR CORPORATION, Aichi, Japan, Shin-ichi Tsuruta, Japan International Research Center for Agricultural Sciences (JIRCAS), Okinawa, Japan, Yoshifumi Terajima, Japan International Research Center for Agricultural Sciences, Ishigaki, Japan, Shotaro Ando, Japan International Research Center for Agricultural Sciences (JIRCAS), Japan and Suchirat Sakuanrungrasirikul, Khon Kaen Field Crops Research Center (KKFCRC), Khon Kaen, Thailand

In order to promote sugarcane breeding, mating between lines with genetic variation is important, and genome analysis is necessary for this purpose. Sugarcane is difficult to analyze because of its large genome size and high polyploidy, and there are few reports on the genetic relationship between sugarcane lines in Thailand and sugarcane lines in the world.

We analyzed 57 sugarcane lines grown in Thailand and 8 wild species from Thailand and about 200 lines from around the world. The analysis was performed by generating a marker and determining the genotype by Genotyping by Random Amplicon Sequencing-Direct (GRAS-Di), and performing principal component analysis based on the genotype of each line. In the GRAS-Di method, sample preparation was performed by PCR amplification using random primers, NGS data analysis and about 20,000 SNPs markers distributed on the genome on average were obtained.

As a result of the principal component analysis of the genotypes of each line, Thailand lines and wild species are separated in the first principal component (1PC), and the genetic relationship between the sugarcane lines of the world and other wild species was shown.

PE0796: Maize, Sorghum, Millet, Sugar Cane, and related

Hybrid Networks to Shed Light on the Roles of Resistance Gene Analogs on the Modulation of Sugarcane Metabolism during the Interaction with Smut

Hugo V. S. Rody, University of São Paulo, Piracicaba, Sao Paulo, Brazil, Renato G. H. Bombardelli, University of Sao Paulo, Piracicaba, Brazil and Claudia B. Monteiro-Vitorello, University of São Paulo, Piracicaba, Brazil

Sugarcane-smut caused by the fungus *Sporisorium scitamineum* is spread worldwide and during severe infections, may result in production losses up to 62%. Although oxidative burst has been reported for the early moments of smut-sugarcane interaction, no *in silico* large-scale investigation of sugarcane immune system has been performed. Herein, we used complex networks to study the roles of predicted resistance gene analogs (RGAs) available at the Sugarcane Orthologs of Resistance Database (SORD) on the modulation of sugarcane metabolism during early interaction with smut. We established hybrid networks for sugarcane using metabolic and text-mining data from its closely related species of *Sorghum bicolor* available at KEGG and String databases, respectively. In addition, we used transcriptomic data from two sugarcane genotypes with contrasting degrees of resistance to smut to unravel metabolism modulation of sugarcane during early interaction with smut. A single glycosyltransferase from N-Glycan biosynthesis pathway was found as creating edges to three RGAs from TM-CC class. Depth-first search analysis using the glycosyltransferase node as a source highlighted genotypes disparities in the modulation of carbohydrate metabolism. Three nodes from N-Glycan biosynthesis pathway were down-regulated in the smut-

resistant sugarcane genotype, whereas no modulation in this pathway was found for the smut-susceptible genotype. Further, the smut-resistant genotype had more nodes from the carbohydrate metabolism pathways having differentially expressed genes attributed. Our results reiterated the importance of glycan and carbohydrate metabolisms during plant-pathogen interactions and indicated the potential roles of RGAs in the transcript expression of smut-resistant sugarcane.

PO0797: Maize, Sorghum, Millet, Sugar Cane, and related

Maize Transcriptional Responses to Corn Leaf Aphid Infestation

Lise Pingault¹, Suresh Varsani¹, Nathan A. Palmer², Gautam Sarath² and Joe Louis¹, (1)University of Nebraska-Lincoln, Lincoln, NE, (2)Wheat, Sorghum, and Forage Research Unit, USDA-ARS, Lincoln, NE

Maize (*Zea mays* L.) is one of the major cereal crops cultivated throughout the world, with the United States producing over 40 percent of the crop annually. Corn leaf aphid [CLA; *Rhopalosiphum maidis* (Fitch)] is an economically important pest of several monocot crops, including maize. In addition to extensive crop damage, CLA also acts as a vector for viruses that cause devastating diseases in maize. Previously, we showed that the maize inbred line Mp708, which was developed by classical plant breeding, provides enhanced resistance to CLA. Feeding by CLA on Mp708 triggers the rapid accumulation of the *maize insect resistance1* (*mir1*) transcripts, which encodes a cysteine protease. In this study, transcript profiling of CLA susceptible (Tx601) and resistant (Mp708) shoots and roots after foliar release of CLA for 24 hours identified several sets of genes that are differentially regulated and may have important functions in *mir1*-dependent defenses against CLA. The underlying mechanisms of local and systemic CLA-induced signaling networks will be discussed.

PE0798: Maize, Sorghum, Millet, Sugar Cane, and related

Biological Nitrogen Fixation in the Mucilage Produced By Aerial Roots of a Wide Variety of Maize and Sorghum Accessions

Jean-Michel Ane¹, Vania CS Pankiewicz¹, Valentina Infante¹, Claudia I Calderon¹, Wilfred Vermerris² and Natalia de Leon¹, (1)University of Wisconsin - Madison, Madison, WI, (2)University of Florida, Gainesville, FL

Biological nitrogen fixation is a promising and sustainable alternative to the use of synthetic fertilizers that has the potential to increase the profitability and sustainability of cereal crop production. A maize landrace from the Sierra Mixe, Mexico, was found to acquire 30-80% of its nitrogen through biological nitrogen fixation. Bacteria are hosted in a mucilage produced by aerial roots after rain. We assessed, in field conditions, 64 accessions from the CIMMYT and the USDA GRIN public collections. We selected 14 elite accessions that developed more than seven nodes with aerial roots thicker than 0.6 cm in diameter and produce copious amounts of mucilage. Those lines are currently evaluated for biological nitrogen fixation and used for breeding the trait into conventional maize cultivars as well as identification of the genes controlling the trait. We showed that the ability to produce abundant aerial roots and mucilage is a trait present in teosinte, the progenitor of maize, but also in other cereal crops, such as sorghum. We identified sorghum accessions that produce large amounts of mucilage after rain and support levels of nitrogenase activity similar to Sierra Mixe maize. The mucilage produced by sorghum aerial roots has a sugar composition, viscosity, and properties similar to the maize one. Using the ¹⁵N dilution technique, we demonstrated that these sorghum accessions acquire about 10% of their nitrogen from the air. Altogether our data indicate that biological nitrogen fixation in the mucilage produced by aerial roots is a conserved trait throughout cereals.

PO0799: Maize, Sorghum, Millet, Sugar Cane, and related

Conserved Transcriptional Responses for Leaf Blight in Sorghum and Maize

Pragya Adhikari, Santiago X Mideros and Tiffany Jamann, University of Illinois at Urbana-Champaign, Urbana, IL

Exserohilum turcicum is an important fungal pathogen of both sorghum and maize, causing sorghum leaf blight and northern corn leaf blight. Because the same pathogen can infect two of the most important grain crops, it is an ideal

pathosystem to study plant-pathogen co-evolution and investigate shared resistance mechanisms between the two plant species. In this study, we performed transcriptional analysis of maize and sorghum in response to maize-specific (compatible maize and incompatible sorghum interaction) and sorghum-specific (compatible sorghum and incompatible maize interaction) *Exserohilum turcicum* isolates at two-time points (24 hours and 72 hours after inoculation). We compared the maize and sorghum transcriptomes using an ortholog-based approach. In response to compatible pathogen isolates at 24 hours and 72 hours after inoculation, 12 genes and 5 genes were conserved between maize and sorghum, respectively. Likewise, in response to the incompatible pathogen isolate at 24 and 72 hours after inoculation, 4 genes and 58 genes were conserved between maize and sorghum, respectively. Of the genes conserved between both crops, many were defense-related, including pathogenesis-related genes, a chitinase, several resistance genes, cytochrome P450 genes, transcription factors related to stress resistance, a hormone-related gene, and a protease inhibitor. In conclusion, we identified several promising candidate genes for resistance to leaf blight in both maize and sorghum that will be valuable to improving resistance in both crop species to this important pathogen.

PE0800: Maize, Sorghum, Millet, Sugar Cane, and related

Genetic Dissection of Drought Tolerance in a Sorghum Backcross Nested Association Mapping Population

Hongxu Dong, Plant Genome Mapping Laboratory, University of Georgia, Athens, GA

With a worldwide water crisis looming, a primary goal of sorghum breeding is to improve drought tolerance. To address this challenge in sorghum, we developed a backcross nested association mapping (BCNAM) population using 12 diverse founder lines crossed with an Ethiopian elite cultivar Teshale. The sorghum BCNAM population was trialed under two natural drought environments and one normal growing environment in Ethiopia. We characterized 1178 BC₁F₄ lines with 4395 single nucleotide polymorphisms (SNPs) and conducted joint linkage (JL) and genome-wide association analyses (GWAS) for nine adaptive traits. Phenotype was less heritable under drought conditions than under normal growing condition. A total of 177 JL quantitative trait loci (QTL) and 267 GWAS hits were detected across the three environments for the nine traits. 159 JL QTLs (89%) in this study had correspondence with QTLs for the same traits from 72 previous studies. Associations detected in JL and GWAS clustered within known stay-green QTL regions and largely overlapped with the differentially expressed genes between stay-green and senescent sorghum genotypes from previous transcriptomic study. Fixation index (F_{ST}) between a subset of 58 lines that failed to survive drought stress (*i.e.* drought-sensitive) and the remaining 1120 plants (*i.e.* drought-tolerant) showed elevated peaks within stay-green QTL regions, where most associations were detected in this study. Flowering regulator such as *Ma6* and drought resistant gene such as *P5CS2* were in proximity to these associations. Interestingly, using the model-selected SNPs that associated with nine traits across three environments, phenotypic prediction accuracies for grain yield were equivalent to genome-wide SNPs and were significantly better than an equivalent number of random SNPs, indicating that these drought-related traits are predictive of sorghum grain yield. These findings validate the BCNAM resource in trait mapping in sorghum and demonstrate the value of NAM design for dissection of adaptive traits.

PO0801: Maize, Sorghum, Millet, Sugar Cane, and related

Identification of *Sorghum bicolor* SSRs (Xgma1 – Xgma6) for Working Group Margaritiferum Determination

John Baggett and Melinda Yerka, University of Nevada, Reno, Reno, NV

Sorghum [*Sorghum bicolor* (L.) Moench] is a drought tolerant C₄ crop with substantial genetic diversity within the species. To better capture this genetic diversity for breeding purposes, a clear understanding of the heterotic groups available is needed. The five major races of cultivated Sorghum (Bicolor, Caudatum, Durra, Guinea, and Kafir) are able to be differentiated by panicle architecture and seed morphology, but currently a reliable tool to differentiate the Guinea working group, Margaritiferum, is lacking. Accurate Margaritiferum accession identification is valuable because the population is known to have substantial uncaptured diversity that can be used for yield improvements. Short sequence repeats (SSRs) are useful tools to determine genetic relatedness as well as the following of closely linked traits through generations of progeny. In this work, SSRs for identifying accessions belonging to the Guinea working group Margaritiferum are discussed. These 6 SSRs (*Xgma1* - *Xgma6*) can be used for determination of Margaritiferum accessions as well as for future molecular marker development with closely linked genes. The long-

term goal of this research is to provide plant breeders with the tools to make significant yield gains in cultivated Sorghum.

PE0802: Maize, Sorghum, Millet, Sugar Cane, and related

Shriveled SEED (SVD), an HMA5 Family P1B-Type Transporter ATPase, Plays a Critical Role in Growth and Development of Sorghum

Dimiru T Tadesse¹, Liang Sun² and Million Tadege¹, (1)Oklahoma State University, Ardmore, OK, (2)Noble Research Institute, Ardmore, OK

The world population is projected to reach 9.6 billion by 2050, putting strong pressure on food security and agricultural productivity. Understanding plant developmental programs will aid in improving productivity by manipulating genetic pathways and identifying molecular markers that can be used in breeding programs for marker assisted selection. The project generally aims to examine environmentally regulated developmental programs with the ultimate goal of identifying key developmental regulators that integrate environmental signals, and understanding the mechanisms of their actions. Specifically, the study examines heavy metal trafficking and its impact on sorghum growth and development. We used a forward genetics approach employing a sorghum deletion mutant population, and identified a gene which controls heavy metal trafficking. I screened total of 1,200M2 fast neutron irradiated mutant lines in the field to identify a severely affected developmental mutant named *shriveled seeds* (*svd*). The *svd* mutants display growth defects throughout development from germination to seed filling, and appear to be particularly sensitive to heavy metals such as copper and manganese. We cloned the *SVD* gene using whole genome sequencing and bioinformatics analysis, and found out that it encodes a heavy metal transporter P1B-type ATPase. This study is focused in understanding the mechanism of *SVD* function and its impact on sorghum development and maturation.

PO0803: Maize, Sorghum, Millet, Sugar Cane, and related

Genomic Basis of Climatic Adaptation in Sorghum Landraces of Ethiopia

Kassahun Bantte Bisetegn¹, Temesgen Matiwos Menamo¹, Andrew K. Borrell², Emma Mace³, David R Jordan⁴, Colleen H Hunt⁵ and Yongfu Tao⁶, (1)Jimma University, Jimma, Ethiopia, (2)University of Queensland, Warwick, Australia, (3)Queensland Alliance for Agriculture and Food Innovation, The University of Queensland, Warwick, QLD, Australia, (4)Queensland Alliance for Agriculture and Food Innovation, The University of Queensland, Warwick, Australia, (5)Department of Agriculture and Fisheries, Queensland Government, Warwick, QLD, Australia, (6)The University of Queensland, Warwick, QLD, Australia

Sorghum is the most variable of all crops grown in Ethiopia, containing tremendous genetic diversity. Exploration of this genetic diversity using modern tools is important to associate traits with genomic regions, effectively utilize germplasm, conserve germplasm collections, and obtain core collections for efficient germplasm management. Therefore, the objective of this study was to better understand the genetic diversity and genome-environment association of 940 Ethiopian sorghum landraces using SNP markers. The Ethiopian sorghum germplasm was a genetically diverse collection comprising 12 subpopulations with high levels of admixture (47%). Redundancy analysis indicated that a larger proportion of SNP variation was explained by agroecology (7%) than geographical location (3%) and agroecology collinear with geographical location (3%). Most of the subpopulations belonged to the durra botanical race and none of the subpopulations belonged to the kafir botanical race. Ethiopian sorghum was distributed throughout most of the agro-ecological zones. Genome-environment association (GEA) studies with altitude, annual temperature and precipitation variables identified a total of 18 significant SNP markers for these environmental variables. These significant environmental SNP markers were co-located with 74 previously identified drought and cold adaptation QTLs. There was a significant enrichment of GEA SNP markers with a priori candidate QTLs for drought and cold adaptation. Based on this result, the Ethiopian sorghum germplasm collection was highly diverse with four sorghum botanical races (except kafir) with adaptations to extreme environments (cool/humid and warm arid) which could be a source of traits for abiotic stresses such as cold and drought adaptation for future breeding programs.

PE0804: Maize, Sorghum, Millet, Sugar Cane, and related

Genome-Wide Prediction, Association, and Gene Network Analysis of Grain Composition in Sorghum

Sirjan Sapkota^{1,2}, J. Lucas Boatwright², Richard E. Boyles^{1,3} and Stephen Kresovich^{1,2}, (1)Department of Plant and Environmental Sciences, Clemson University, Clemson, SC, (2)Advanced Plant Technology Program, Clemson University, Clemson, SC, (3)Pee Dee Research and Education Center, Clemson University, Darlington, SC

Starch and protein are two of the most important constituents of grain contributing to most of human and animal caloric needs. The genetic control of starch and protein composition are complex and not completely understood. While genome-wide prediction has been routinely studied for grain yield, little is done in terms of application of predictions for grain composition. Here we present our genotype-phenotype association and prediction study for starch, protein and gross energy using 224,007 SNPs in a sorghum diversity panel with 389 individuals. We didn't find any significant difference in predictive ability between various Bayesian models with different priors. On average the predictive ability was 0.6 for starch, 0.45 for protein, and 0.58 for gross energy. Using multivariate linear mixed model for starch and protein we were able to identify significant associations at genomic regions in chromosomes four and eight that were not significant using univariate model for starch or protein. A total of 13 genes within linkage disequilibrium of the associated regions had high confidence (0.7) first interactors using STRING. The gene network analysis of those 13 genes and their first interactors showed significant enrichment for various biochemical pathways including sucrose and starch biosynthesis and nitrogen metabolism. Our results provide new insights into application of multivariate approach in statistical learning. Furthermore, the genes and pathways identified could be crucial in understanding genetic mechanisms in source-sink dynamics during grain filling.

PO0805: Maize, Sorghum, Millet, Sugar Cane, and related

QTL Analyses for Hybrid Vigor: A Case Study of Sorghum F₁ Variety "Tentak"

Shumpei Hashimoto^{1,2}, Satoko Araki-Nakamura¹, Tatsuro Wake^{1,2}, Miki Yamaguchi^{1,2}, Kozue Shinohara-Ohmae¹, Kotaro Miura³, Shigemitsu Kasuga⁴ and Takashi Sazuka¹, (1)Bioscience and Biotechnology Center, Nagoya University, Nagoya, Japan, (2)Graduate School of Bioagricultural Sciences, Nagoya University, Nagoya, Japan, (3)Faculty of Bioscience and Biotechnology, Fukui Prefectural University, Fukui city, Japan, (4)Faculty of Agriculture, Education and Research Center of Alpine Field Science (AFC), Shinshu University, Kamiina, Japan

F₁ hybrids are widely cultivated because they have superior performance in agronomical traits compared with inbred lines in sorghum. A typical case is the F₁ hybrid variety "Tentak", which is a Japanese tall cultivar (~4 m tall), although the both parents are short (~1.5 m tall). This phenomenon has been recognized as a typical example of hybrid vigor. Here, we report the QTL analyses of biomass traits of Tentak.

We studied flowering time, and culm length using an F₂ population of Tentak by QTL analyses. The analysis of flowering time detected two QTLs. The pollen parent allele of *qFD-1*, and the seed parent allele of *qFD-6* fastened the flowering time. The analysis of culm length detected four QTLs. The seed parent alleles of *qCL-6* and *qCL-7a*, and pollen parent alleles of *qCL-7b* and *qCL-9* shortened the culm length. All six alleles were recessive. If we describe the responsible genes *qFD-1*, *qFD-6*, *qCL-6*, *qCL-7a*, *qCL-7b*, and *qCL-9* as *A*, *B*, *C*, *D*, *E*, *F*, the follows were suggested; the seed parent is *AAbbccddEEFF*, the pollen parent is *aaBBCCDDeeff*, and the F₁-hybrid "Tentak" is *AaBbCcDdEeFf*. These results also suggest that the hybrid vigor of "Tentak" could be mainly explained by the dominant theory. Further analysis revealed that the candidate genes of *A*, *B*, *E*, and *F* are *SbPhyB*, *SbGhd7*, *Dw3*, and *Dw1*, respectively.

This work was partially supported by JST-Mirai Program Grant Number JPMJMI17EG, Nagoya University-RIKEN Science, Technology and Innovation Hub collaboration project, and the Sasakawa Scientific Research Grant from The Japan Science Society.

PE0806: Maize, Sorghum, Millet, Sugar Cane, and related

Allelism Test and Morphological Characterization of Sorghum Male Sterile Mutants

Junping Chen, USDA-ARS, Lubbock, TX, Haydee Laza, Texas Tech, Lubbock, TX and Zhanguo Xin, USDA ARS, Lubbock, TX

Identification of genes involved in anther development and understanding of the functions and regulations of these genes in anther and pollen development is essential to develop male sterile lines in hybrid breeding. In this study, we isolated a large collection of nuclear male sterile (*nms*) mutants in sorghum [*Sorghum bicolor* (L.) Moench] from an ethyl methane sulfonate–mutagenized (EMS) mutant population. To effectively identify the causal mutations in these mutants, we performed genetic complementation tests with the 4 previously known *nms* lines and identified new alleles for these known *nms* lines. We also identified many novel genetic loci in which, a number of them have multiple independent alleles, making the identification of the underlying gene mutation much easier. Furthermore, we performed morphological and cytological characterization of the sorghum *nms* mutant lines and classified them into several groups based on their morphological differences in anther development. We have also successfully identified the causal gene mutations represented by the *ms9* mutant in the first novel *nms* group. This vast collection of sorghum *nms* mutants, together with the causal genes identified, provide new genetic tools for studying pollen development and designing new efficient strategies to breed sorghum hybrids.

PO0807: Maize, Sorghum, Millet, Sugar Cane, and related

Dicer-like 5 Deficiency Confers Temperature-Sensitive Male Sterility in Maize

Chong Teng¹, Han Zhang², Reza Hammond³, Kun Huang⁴, Blake Meyers¹ and Virginia Walbot²,
(1)Donald Danforth Plant Science Center, St. Louis, MO, (2)Stanford University, Stanford, CA,
(3)Delaware Biotechnology Institute, Newark, DE, (4)University of Delaware, Newark, DE

Small RNAs play important roles during plant development by regulating transcript levels of target mRNAs, maintaining genome integrity, and reinforcing DNA methylation. *Dicer-like 5* (*Dcl5*) is proposed to be responsible for precise slicing in many monocots to generate diverse 24-nt phased, secondary small interfering RNAs (phasiRNAs), which are exceptionally abundant in meiotic anthers of diverse flowering plants. The importance and functions of these phasiRNAs remain unclear. Here, we characterized several mutants of *dcl5*, including alleles generated by the Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) *Cas9* system and a transposon-disrupted allele. We report that *dcl5* mutants have few or no 24-nt phasiRNAs, develop short anthers and defective tapetal cells, and exhibit temperature-sensitive male fertility. We propose that DCL5 and 24-nt phasiRNAs are critical for fertility under growth regimes for optimal yield.

PE0808: Maize, Sorghum, Millet, Sugar Cane, and related

Discovery of Sorghum Haploid Inducer Lines

Tanveer Hussain, Corteva Agriscience, Johnston, IA

Sorghum, *Sorghum bicolor* L. $2n = 2x = 20$, is the 5th most important cereal grain crop after corn, wheat, rice and pearl millet in the world. Conventional sorghum breeding relies on multiple generations of self-pollination to achieve the adequate levels of homozygosity for hybrid evaluation, which adds several years and great cost to the breeding process. Currently no sorghum doubled haploid (DH) system exists in the industry. Corteva Agriscience is the leader to develop such a system for sorghum breeding. We have discovered two sorghum haploid inducer lines, **SMHI01** and **SMHI02**. The discovery of haploid inducer was the biggest hurdle in bringing the DH technology to sorghum.

The first of its kind, these inducer lines would enable the creation of doubled haploid sorghum, which is the first step towards a revolutionary change and significantly accelerating the sorghum breeding.

PO0809: Maize, Sorghum, Millet, Sugar Cane, and related

Time-Series GWAS Using High-Throughput Phenotyping and Functional Principal Components Analysis in Sorghum

Chenyong Miao, University of Nebraska Lincoln, Lincoln, NE, **Alejandro Pages**, Department of Computer Science, University of Nebraska-Lincoln, Lincoln, NE, **Yuhang Xu**, Bowling Green State University, Bowling Green, OH and **James C Schnable**, Department of Agronomy and Horticulture, University of Nebraska-Lincoln, Lincoln, NE

Compared to genomics and genotyping, until recently approaches to phenotyping had remained reasonably stable for quite some time. The role of collecting trait data as the primary bottleneck in quantitative genetics has spurred an interest in new high throughput approaches to plant phenotyping, including image-based phenotyping. However, several challenges remain including the extraction of numerical traits from images and developing statistical approaches to linking genotype to traits which change dynamically over time. In this study, we use a high-throughput plant phenotyping facility to capture the sorghum association panel (SAP), a set of ~400 diverse sorghum lines, during the transition from vegetative to reproductive development. After performing semantic segmentation using machine learning methods to identify sorghum organs in each hyperspectral image, plant height were extracted for each observed time point and nonparametric regression was employed to estimate the growth curve for each individual. Functional principal components analysis (FPCA) was used to extract the main patterns of the growth curves. The first pattern represents the overall height changes and the second one represents the height changes before and after the panicle emergence. The corresponding two principal components explain over 95% variance of the growth curves for all the individuals. Using the first two principal components as the phenotype to run GWAS, dwarf genes which are related to the growth patterns were identified. Our results show the advantages of using time-series GWAS to identify causal variants and suggest that the FPCA is a robust tool to map dynamic traits in association studies.

PE0810: Maize, Sorghum, Millet, Sugar Cane, and related

Mutation Breeding Creates Desired Traits for African Sorghum –Semi-Dwarf and Early Maturing

Norman Warthmann, Abdelbagi M. Ali Ghanim, Adel Ali and Ivan Ingelbrecht, Plant Breeding and Genetics Laboratory, FAO/IAEA Joint Division, International Atomic Energy Agency (IAEA), Seibersdorf, Vienna, Austria

Wad Ahmed is a Sorghum variety popular among farmers in Sudan, except that it matures slightly late and is tall, making it prone to yield losses caused by terminal drought and lodging. Gamma-irradiation of seeds followed by breeding work was undertaken to quickly and cost-effectively obtain early-maturing and semi-dwarf mutants and derived populations (M_6). By high-throughput short-read sequencing we compared the mutants to their progenitors and genotyped large F_2 -populations ($M_6BC_1F_2$) for genetic mapping. We are developing molecular markers to facilitate the use of these new traits in Sorghum breeding programs in Member States.

PO0811: Maize, Sorghum, Millet, Sugar Cane, and related

Composition and Genomics of Surface Chemicals on Grain of *Sorghum bicolor* and Related Grasses

Lucas Busta, University of Nebraska Lincoln, Lincoln, NE

Sorghum bicolor is a multi-use crop with exceptional water- and nitrogen-use efficiency whose expanded use will likely decrease water and fertilizer consumption and create a more sustainable agricultural system. However, sorghum is generally not economically competitive with maize when irrigation is available, so sorghum harvest value needs to be increased. Natural waxes cover aerial plant surfaces where they protect against dry climates and increase water-use efficiency. On sorghum, waxes accumulate to levels higher than nearly all other plant species and sorghum kernel waxes are an emerging industrial bioproduct that could replace ubiquitous carnauba wax as sorghum waxes can be extracted during grain processing. The long-term goal of this project is to combine wax biochemistry with genomics and bioinformatics to identify genes controlling sorghum wax deposition. The goal of the present work is to perform a detailed chemical characterization of waxes on sorghum grain and to uncover genes controlling the chemical profile of the kernel surface.

We first extracted and fractionated sorghum grain surface chemicals into compound classes (primary alcohols, secondary alcohols, and aldehydes) and compared their melting properties against those of carnauba. The primary alcohol fraction from sorghum waxes had melting properties similar to those of carnauba wax. We next used the chemical composition information to guide transcriptome mining in public sorghum RNA-Seq datasets, resulting in

the identification of 14 genes potentially involved in sorghum wax biosynthesis. We are currently working on functionally testing these genes using heterologous expression. The characteristics (expression, presence/absence, primary sequence) of sorghum wax biosynthesis genes are also being compared with related genes from other grass species, all in the context each grasses' grain wax profile.

PE0812: Maize, Sorghum, Millet, Sugar Cane, and related

ARG2 Confers a Complete Race-Specific Resistance to Sorghum Anthracnose Disease

Demeke Mewa Bayable, Purdue University, West Lafayette, IN

Anthracnose disease caused by the fungal pathogen *Colletotrichum sublineolum* is a widespread and economically important disease of sorghum. Genetic resistance to this disease is the only economically feasible disease control strategy in most sorghum producing regions. We explored some natural variation in sorghum and genetically defined the *ANTHRACNOSE RESISTANCE GENE 2* available in a resistant sorghum line. The line was selected based on a clear-cut and easily distinguishable host-response to the pathogen strain during the initial genetic evaluation to identify the resistant genotypes. Subsequently, to identify the *ARG2* gene, a genetic mapping population was developed by crossing the resistant line to TAM428 which is highly susceptible to most strains of the pathogen. Genetic analyses of the host-response at F1 and F2 generations suggested that the resistant line carries a dominant resistance gene to some but not all strains of *C. sublineolum*. Subsequently, bulk-segregant analysis using whole-genome re-sequencing (BSA-Seq) of F2 plants mapped the *ARG2* locus to the near-proximal end of the first arm of chromosome05. This *ARG2* candidate genomic region was verified and substantially narrowed using recombination analysis and the annotated biological function. A comparative genomic analysis followed by the examination of gene expression in adjacent duplicate R-genes of the locus and base-sequence analysis between the parental lines strongly identified the likely candidate *ARG2* gene. The candidate gene is validated by the identification of independent mutant allele in the *ARG2* locus which is susceptible to the same strains of the fungus. In sum, we present evidence for the identification of an anthracnose resistance gene and the molecular characterization of the *ARG2* is underway.

PO0813: Maize, Sorghum, Millet, Sugar Cane, and related

Comparison of 16 Wild and Domesticated Sorghum Genomes Reveals Extensive Variation in the Sorghum Pan-Genome

Yongfu Tao¹, Hong Luo², Jiabao Xu³, Alan W Cruickshank⁴, Fei Teng³, Tracey Shatte⁴, Xiaoyuan Wu², Yuanming Liu², Haichun Jing², David Jordan⁵ and Emma Mace⁵, (1)The University of Queensland, Warwick, QLD, Australia, (2)Institute of Botany, Chinese Academy of Sciences, Beijing, China, (3)BGI, SHENZHEN, China, (4)Department of Agriculture and Fisheries, Queensland Government, Warwick, QLD, Australia, (5)Queensland Alliance for Agriculture and Food Innovation, The University of Queensland, Warwick, QLD, Australia

Sorghum provides a major source of food, feed, fibre and biofuel. Cultivated sorghum and its inter-fertile wild relatives have a broad geographical distribution and constitute the primary gene pool for sorghum. The rich diversity within this resource is the raw material upon which human selection act during domestication and modern breeding to reshape its phenology and productivity to meet human needs. To better understand diversity in the sorghum primary gene pool, we have constructed the sorghum pan-genome using 16 de novo genome assemblies. We de novo assembled 13 sorghum genomes including cultivated sorghum and its wild relatives via a hybrid sequencing strategy combining Illumina short reads and PacBio long reads. Analysis of the sorghum pan-genome was conducted using these 16 assembled genomes, which revealed a sorghum pan-genome consisting of 61,564 predicted gene families. Of these, approximately 30% (18,141 gene families) were present in ≥ 15 sorghum genomes, and defined as core genes, 33% (20,439) were present in 2-15 sorghum genomes, and defined as shell genes, and 37% (22,984) were present in only one sorghum genome, and defined as cloud genes. Compared to core genes, shell genes were found to be shorter with less exons, but with higher SNP density and higher non-synonymous/synonymous substitution ratios, indicating that shell genes are less functionally conserved. Comparisons with the reference genome identified 1.23-5.31 million SNPs and 20,386-148,669 PAVs in each comparison. Genes affected by PAVs were enriched with defence response functions. Selection signals on PAVs during sorghum domestication were identified and associations between PAVs and agronomical traits were detected.

This study presents the first pan-genome analysis in sorghum with thorough investigation on PAVs. Results from this study will be a good resource for genetic and genomic analysis in sorghum and other crops.

PE0814: Maize, Sorghum, Millet, Sugar Cane, and related

Phenotypic and Phytochemical Analysis of Black Sorghum (*Sorghum bicolor* L.) Pericarp in Response to Light Quality

Lauren N. Fedenia, Texas A&M Department of Horticulture, College Station, TX, Patricia E. Klein, Department of Horticultural Sciences, Texas A&M University, College Station, TX, Robert R. Klein, USDA-ARS Crop Germplasm Research Unit, College Station, TX, Linda Dykes, USDA-ARS, Fargo, ND and William L. Rooney, Texas A&M University, College Station, TX

The black pericarp trait of sorghum [*Sorghum bicolor* (L.) Moench] has notable applications in the specialty food, nutraceutical, and medical industries. The black appearance of this grain is associated with the production of 3-deoxyanthocyanidins (3-DOA), a group of phytochemicals valued for their cytotoxicity to human cancer cells and high molecular stability for practical use as natural food colorants and antioxidant food additives. Prior observations revealed that black pericarp phenotype is not fully expressed in all environments leading to the hypothesis that light quality and/or duration (photoperiod) affects expression of the trait. The objective of this study was to examine how light quality effects the development of the black pericarp trait by quantifying the changes in 3-DOA accumulation under different light quality regimes. A sorghum genotype with uniformly black-pericarp sorghum was grown in chambers under selective penetration of UV-B, UV-A, photosynthetically active light (PAR), or no light penetrance. As a control, a red-pericarp sorghum was exposed to the same light quality regimes. Using ultra-performance liquid chromatography coupled with photodiode array detection (UPLC-PDA), we examined the changes in 3-DOA biosynthesis under the different light treatments while concomitantly measuring reactive oxygen species activity. These studies revealed UV-B light is obligatory for full development of black pericarp and 3-DOA biosynthesis. Finally, the involvement of reactive oxygen species in the development of the black pericarp trait and the biosynthesis of 3-DOAs will be discussed.

PO0815: Maize, Sorghum, Millet, Sugar Cane, and related

Draft Assembly of the *Vossia cuspidata* Genome and Comparisons to Maize

Sontosh K. Deb¹, Michelle L. Lewis¹, M. Cinta Romay², Evan Rees³, Baoxing Song³, Nicholas K. Lepak⁴, Remy Pasquet⁵, Edward S. Buckler² and Michael McKain¹, (1)The University of Alabama, Tuscaloosa, AL, (2)Institute for Genomic Diversity, Cornell University, Ithaca, NY, (3)Cornell University, Ithaca, NY, (4)USDA-ARS, Ithaca, NY, (5)DIADÉ, Montpellier, France

Polyploidy is prevalent in the tribe Andropogoneae (Panicoideae; Poaceae) with over 30 identified, independent events. Andropogoneae includes many ecologically and economically important species like maize, sorghum, and sugarcane. Maize (*Zea mays* ssp. *mays*) is a suspected ancient allopolyploid having originated ~10 mya after the merging and doubling of two unknown ancestral genomes and prior to the diversification of the *Zea* and *Tripsacum* genera. Despite sharing the same polyploid event, *Zea* (n=10) and *Tripsacum* (n=18) experienced distinct chromosomal rearrangements demonstrating the unique trajectory that divergent lineages can take during the diploidization process. Recently, *Vossia cuspidata* (hippo grass) and some other close relatives in Andropogoneae were suggested to form a clade of putative extant relatives of one the maize diploid progenitors. Here we present the initial genome assembly of *Vossia cuspidata* and conduct comparative phylogenomic analyses with maize, sorghum, and *Setaria italica* to test the hypotheses of *Vossia* being relative of a maize diploid progenitor. We also investigated the signal of homoeologous exchanges with replacement between maize subgenomes. Evidence from these analyses suggests that, *Vossia* is more closely related to one of the maize diploid progenitors and that homoeologous exchanges potentially obscure the ancestry of maize subgenomes. Further study involving the complete assembly of *Vossia cuspidata* genome, other *Zea* and *Tripsacum* genome, and expansion of our analyses to microsyntenic blocks from annotated genes will provide a more detailed understanding of rearrangement and homoeologous exchanges post polyploidy within the maize genome.

PE0816: Maize, Sorghum, Millet, Sugar Cane, and related

Differential Gene Expression Studies for Nitrogen Metabolism in Pearl Millet Genotypes Contrasting in Low Soil Nitrogen

Pusuluri Madhu, International Crops Research Institute for the Semi-Arid Tropics, Hyderabad, India and Vijaya Lakshmi Pujarula, Rakesh Kumar Srivastava, Rajeev Gupta

Changing climate, depleting natural resources, and loss of soil fertility is threatening to reduce crop productivity in the dry and marginal environment globally. Pearl millet (*Pennisetum glaucum* (L.) R. Br.) being iron & zinc-rich is the sixth most important cereal crop in the world in terms of area, and is well adapted to marginal production systems characterized by low rainfall and high temperatures. Climate change due to the excess use of Nitrogen and release of its related gases from the agricultural lands need to be regulated by selecting the genotypes with high yield coupled with low nitrogen application. Even though most of the pearl millet production environments have low nitrogen (N) in the root zone soil strata, with the priority of the low input sustainable and environment friendly agriculture, developing pearl millet genotypes with high Nitrogen Use Efficiency (NUE) is highly remunerative and rewarding. To achieve these targets one should have a detailed understanding of the phenotypic trait variations for the NUE trait among the diversified panel of Pearl millet germplasm and indeed should have a detailed understanding of the genes governing high NUE. Aiming to identify the responsible genes and its components in imparting high yield under low N inputs, it was planned to screen and identify natural variations for Nitrogen Use Efficiency in a diversified pearl millet inbred germplasm association panel (PMiGAP) by ICRISAT under CINTRIN (Cambridge India Network for Translational Research in Nitrogen) project. Relying on the basic phenotypic studies two sets of contrasting genotypes towards nitrogen tolerance and nitrogen sensitive were selected and N stimulated gene expression profile was conducted by utilizing the transcripts derived from flag leaf and seeds. Gene expression studies were conducted for a total of six major genes involved in N metabolism viz. *PgNR* (Nitrate reductase), *PgNiR* (Nitrite reductase), *PgGS* (Glutamine synthetase), *PgGOGAT* (Glutamine oxoglutarate aminotransferase), *PgAlaAT* (Alanine aminotransferase) and *PgAS* (Asparagine synthetase). The results have given the preliminary idea about the low nitrogen tolerance at the molecular level; almost all the genes examined were upregulated in Jakhrana under both N0 and N100 levels whereas in RIB the expression of these crucial genes was downregulated at N0 and showed a trend towards upward regulation when the plants were induced with N100. This examination is ideal for the identification of contrasting lines for high and low NUE and also specifically explains the crucial role of N metabolism genes in imparting the tolerance towards low-nitrogen conditions.

PO0817: Maize, Sorghum, Millet, Sugar Cane, and related

Genetic Analysis of Arabinoxylan Biosynthesis in Maize Seed and Impact on the Human Gut Microbiome

Nathaniel Korth¹, James C Schnable² and Andrew Benson¹, (1)University of Nebraska-Lincoln, Lincoln, NE, (2)Center for Plant Science Innovation, University of Nebraska-Lincoln, Lincoln, NE

Microorganisms that inhabit the human gut are essential for life and have been implicated in various health and disease states. Diet is thought to be the greatest driver of configuring the species composition of the human gut microbiome. Despite the connection of diet and the microbiome, crop breeding and improvement strategies have not traditionally been focused on nutritional outcomes that impact the gut microbiome. Arabinoxylan (AOX) is a hemicellulose plant cell wall material that varies structurally between different plant species, some of these structures are utilized by gut microbes post digestion. However, the relationship between the AOX fiber components and microbes that degrade them is largely unknown. We have identified maize lines with transposon insertions that alter genes implicated in different aspects of AOX biosynthesis including xylan backbone elongation, and crosslinking of xylan molecules by ferulic acid and arabinose. Target genes were chosen based on function and expression in the kernels and were identified from the Uniform Mu transposon mutagenesis population developed at the University of Florida. Subsets of first-generation seed were processed in a simulated human digestion. The digested product was introduced individually into *in vitro* microbiome reactions, spanning multiple human subjects. The microbial composition of the microbiome was analyzed by 16S rRNA sequencing. Preliminary results indicate mutations in xylan elongation and ferulic acid crosslinking impacts growth of specific groups of gut bacteria. Characterization of microbial responses to structurally distinct AOX mutants will provide a comprehensive view of the how these complex molecules can influence the microbiome and health.

PO0819: Maize, Sorghum, Millet, Sugar Cane, and related

Co-Evolution of Crops, Birds, and Farmers in African Agroecosystem Linked by Secondary Metabolites

Yuye Wu¹, Tingting Guo², Qi Mu², Jinyu Wang², Xin Li², Yun Wu², Bin Tian³, MingLi Wang⁴, Guihua Bai⁵, Ramasamy Perumal¹, Harold N. Trick³, Scott Bean⁶, Ismail Dweikat⁷, Mitchell Tuinstra⁸, Geoffrey P. Morris¹, Tesfaye T. Tesso¹, Jianming Yu² and **Xianran Li²**, (1)Kansas State University, Manhattan, KS, (2)Iowa State University, Ames, IA, (3)Kansas State University, Department of Plant Pathology, Manhattan, KS, (4)USDA-ARS, Griffin, GA, (5)USDA, Hard Winter Wheat Genetics Research Unit, Manhattan, KS, (6)USDA-ARS, Manhattan, KS, (7)University of Nebraska-Lincoln, Lincoln, NE, (8)Purdue University, WEST LAFAYETTE, IN

Domestication is an evolutionary process of humans selecting desired characterizes from wild progenitors to fit human needs and local environments. All three components, human (domesticator), crop (domesticate), and environment are essential for domestication to take place.

Among major cereals domesticated as staple foods, only sorghum (domesticated in Africa) has a high proportion of cultivars with condensed tannins, which can trigger bitter perception in animals by binding to type 2 taste receptors (TAS2Rs). We identified a pair of duplicate recessive genes (*Tan1* and *Tan2*) underlying the presence of tannins in sorghum grains. Three loss-of-function alleles from each gene were identified in non-tannin sorghum desired as a palatable food. Following the serendipitous observation that condensed tannins effectively prevented sparrows from consuming sorghum grain, we uncovered parallel geographic distributions between tannin sorghum and the red-billed quelea bird, supporting the role of tannins in fighting against this major herbivore threat in Africa. Association between geographic distributions of human *TAS2R* variants and tannin sorghum suggested that people in areas where tannin sorghum are predominantly grown are more likely to carry the non-taster alleles in *TAS2Rs*, presumably better tolerance to the bitter taste from tannins.

With the uncovered genetic evidence and the parallel distributions, condensed tannins probably played important roles in African agroecosystem. According to local environments and human taste sensitivity, the balance between natural and artificial selection resulted in contrasting directions: condensed tannins were selected for to reduce herbivore damages in East and South Africa and were selected against to produce palatable food in West Africa. Crop domestication is an intricate process of dynamically balancing the interactions among the plant-human-environment triangle.

PE0820: Oilseeds, Sunflower, and related

Cle-Family Peptide Signaling Regulates Diverse Floral Morphologies across Plant Taxa

Daniel Jones, Amala John and Zachary L. Nimchuk, University of North Carolina, Chapel Hill, NC

Understanding how conserved signaling pathways can coordinate diverse developmental forms is a central question in biology. Signaling via CLE (CLV3/EMBRYO-SURROUNDING REGION-related) family peptides is a conserved cell-cell communication mechanism regulating stem cell identity, division-plane orientation and organogenesis across highly divergent plant taxa. In this study, we have identified and compared CLE-signaling components that regulate inflorescence development in two model systems that have very distinct inflorescence morphologies, *Arabidopsis thaliana* (Arabidopsis) and *Helianthus annuus* (sunflower). Species in the sunflower family (Asteraceae) all have a compact inflorescence known as a capitulum, an evolutionary innovation thought to be crucial for the Asteraceae's large expansion (~23,000 species) and success (global distribution). It has been hypothesized that the capitulum evolved from a diffuse inflorescence type (a more common morphology among flowering plants), in a process that would require widening of the shoot meristem via stem cell proliferation as well as the suppression of internode elongation between flowers. Through the use of genetics, comparative genomics and transcriptomics; we have identified key differences in CLE signaling components that may account for the development of a capitulum instead of a more diffuse raceme (like in Arabidopsis). These results may provide evidence as to how the capitulum first evolved and can lend insight into understanding how CLE signaling affects overall plant morphology.

PO0821: Oilseeds, Sunflower, and related

Unraveling the Sclerotinia Basal Stalk Rot Resistance Derived from Wild *Helianthus argophyllus* Using a High-Density SNP Linkage Map

Zahirul I. Talukder¹, William Underwood², Christopher G. Misar², Gerald J. Seiler², Xuehui Li¹, Xiwen Cai³ and Lili Qi², (1)North Dakota State University, Fargo, ND, (2)USDA-ARS, Edward T. Schafer Agricultural Research Center, Fargo, ND, (3)Department of Plant Science, North Dakota State University, Fargo, ND

Basal stalk rot (BSR), caused by the fungus *Sclerotinia sclerotiorum*, is a serious disease of sunflower (*Helianthus annuus* L.) in the humid temperate growing areas of the world. The genetics of BSR resistance is quantitative and conditioned by many small effect genes. Our objective was to dissect the BSR resistance introduced from the wild annual species *H. argophyllus* using a quantitative trait loci (QTL) mapping approach. An advanced backcross population (AB-QTL) with 134 lines derived from the cross of HA 89 with a *H. argophyllus* accession, PI 494573, was evaluated for BSR resistance in the field during 2017-2018, and in the greenhouse in 2019. Highly significant genetic variations ($p < 0.001$) were observed for BSR disease incidence (DI) in both field seasons, and disease rating and area under the disease progress curve in the greenhouse. The parents and the AB-QTL population were genotyped using genotyping-by-sequencing. A genetic linkage map spanning 2045.14 cM was constructed using 3,112 SNP markers mapped on 17 sunflower chromosomes. A total of 17 QTL associated with BSR resistance were detected on 12 chromosomes, each explaining a phenotypic variation ranging from 5.5% to 19.7%. Of the 17 QTL, six QTL were detected for BSR DI measured in the field, six were detected for traits measured in the greenhouse, and five were detected from both field and greenhouse tests. Twelve of the 17 QTL had favorable alleles from the *H. argophyllus* parent conferring increased BSR resistance.

PE0822: Oilseeds, Sunflower, and related

Association Studies for the Cultivated Sunflower Oil Improvement

Alina Chernova, Skolkovo Institute of Science and Technology, Moscow, Russian Federation

Cultivated sunflower is one of the key plants used by humans. It is an important oilseed crop that was domesticated from the wild sunflower in North America approximately 4000 years ago. Now sunflower is mainly planted for seed oil. The selection of hybrids with changed oil properties is one of the basic directions in oilseed crops hybrid breeding. Full sunflower genome assembly released by Badouin et al. made good possibilities for large-scale genome-wide association studies (GWAS) between genotype and phenotypical characteristics. In this study, we perform high-throughput genotyping (GBS-sequencing) and lipidomic phenotyping on 600 inbred sunflower lines from the Russian sunflower germplasm collection. For lipid profiling, we use ultra-high-performance liquid chromatography coupled with mass spectrometry (UPLC-MS). UPLC-MS is a very powerful tool for lipidomics, which allows the simultaneous profiling of several hundred different lipid species extracted from a single plant sample and to detect minor oil compounds that were not properly studied yet. Here we combine NGS based genotyping with high-performance phenotyping technology and show advantages of this approach for agricultural proposes. We have estimated Russian sunflower germplasm variability and compared it with wild-type sunflower lines and other cultivated lines. Then the lipidome analysis was performed and 1000 lipid molecules were detected. Few lipidome phenotypes were selected for GWAS. Significant associations between major and minor fatty acids and SNPs were identified. Our results extend current knowledge of sunflower genetics and metabolism and give new insights on the development of new approaches in oil-seed crop genomic selection.

PO0823: Oilseeds, Sunflower, and related

Signals Regulating Sunflower Response to Weeds

David P. Horvath¹, Rand Bieri², Wun S. Chao¹ and James V. Anderson¹, (1)USDA-ARS, Fargo, ND, (2)University of Wisconsin - Superior, Superior, WI

Above and belowground weed-generated signals are perceived by sunflowers and induce developmental changes in the crop that reduce yield. We investigated the impact on growth and development as well as transcriptome responses in sunflower roots and leaves exposed to either below-ground, above-ground, or a combination of the two signal sources produced by canola growing in the same pots as the sunflower plants. To isolate the impact of below-ground signals, a single sunflower plant was grown under greenhouse conditions and surrounded by 4 canola plants

growing in the same pot. The above-ground canola-generated signal was blocked by an opaque barrier placed between the sunflower plants and the canola plants. Likewise, to isolate the effects of the above-ground signals, the below-ground signals were locked by planting the canola in cone-tainers that prevented direct root-to-root contact between the two species. Controls where both the above and below-ground signals were blocked or not blocked were included in the experimental design. A complete random block experimental design with 6 replicates was run twice, and phenological measures of growth and RNA collected from roots and leaves of the sunflower were collected and analyzed. Plants were fertilized weekly and watered daily and the sunflower plants rapidly overtopped and were not directly shaded by the canola. The results indicated that the below-ground signals had a greater impact than above-ground signals, and RNAseq analysis identified differentially expressed genes impacted by both above- and below-ground signals.

PE0824: Oilseeds, Sunflower, and related

Association Studies Regarding the Fertility Restorer Gene *Rf1* in Sunflower

Renate Horn¹, Aleksandra Radanovic², Lena Fuhrmann¹, Yves Sprycha¹, Sonia Hamrit³, Milan Jockovic², Dragana Miladinovic² and Constantin Jansen³, (1)University of Rostock, Rostock, Germany, (2)Industrial Crops Department, Institute of Field and Vegetable Crops, Novi Sad, Serbia, (3)Strube Research GmbH & Co. KG, Söllingen, Germany

Cultivation of sunflower is based on hybrid breeding using world-wide a single cytoplasmic male sterility (CMS) source, named PET1-cytoplasma. This CMS source was obtained by crossing *H. petiolaris* with *H. annuus*. For fertility restoration two restorer genes are known, of which one, the *Rf2* gene, is present in most lines. The restorer gene *Rf1* is the major gene responsible for restoration of pollen fertility in sunflower hybrids. The restorer gene *Rf1* has been mapped to the linkage group 13 of the sunflower reference genetic map. Using sequences of known markers linked to the restorer gene as well as BAC end sequences we could physically localize the *Rf1* gene to two potential locations, covering 30 Mb and 3.9 Mb, on LG13 in the HanXRQ genome sequence. In these two regions nine potential candidate genes had been annotated in the sunflower genome sequence coding for seven pentatricopeptide repeat (PPR) genes, a probable aldehyde dehydrogenase 22A1 and a poly(A) polymerase 3. Performing amplicon targeted next generation sequencing for these genes 210 Single Nucleotide Polymorphisms (SNPs) and 67 Insertion/Deletions (InDels) could be identified between the investigated 27 maintainer and 32 restorer lines. Ten SNPs located in three of the PPR genes showed significant association with fertility restoration and one of these genes might be the restorer gene *Rf1*. The genetic differences can now be further analyzed regarding the effects on the genes and can be used to develop SNP-based markers for hybrid breeding.

PO0825: Oilseeds, Sunflower, and related

Genetic Control of Stomatal Morphology and Vein Architecture in a Diversity Panel of Cultivated Sunflower

Ashley Marie Earley, Andries A. Temme, Lisa A Donovan and John M. Burke, University of Georgia, Athens, GA

Drought is a major agricultural challenge that limits plant growth and productivity worldwide. As climate change worsens, droughts are expected to increase in frequency and severity. Stomata and veins both play an essential role in plant growth by facilitating transpiration and movement of water throughout the leaves which may be important for how plants respond to drought stress. This project aims to determine the amount of variation present in stomatal density, stomatal size, and vein density of major and minor veins across 239 genotypes of cultivated *Helianthus annuus* and to determine correlations between these traits. Leaves have been collected for 4 replicates of the 239 genotypes. Stomatal impressions were taken for the top and bottom of the most recent fully expanded leaf from each plant. In addition, leaf samples were preserved in FAA and are being cleared and stained for imaging of major and minor veins. This allowed us to collect data on stomata size and density as well as vein density for the major and minor veins. Results indicate that there is a greater than three-fold difference in values for stomatal density along with substantial variation in stomatal size and vein density. Genome wide association studies (GWAS) has been used to determine which regions of the genome are associated with these traits and which traits co-localize to the same regions of the genome. Future experiments will be conducted to see if any of these traits predict performance under drought stress.

PE0826: Oilseeds, Sunflower, and related

Drought Stress Response in Cultivated Sunflower (*Helianthus annuus* L.)

Garrett M. Janzen¹, Emily L. Dittmar² and John M. Burke², (1)University of Georgia, Athens, IA, (2)University of Georgia, Athens, GA

Drought stress is a common concern for many cultivated crops. However, the effects of drought stress on a crop may vary depending on drought intensity and duration and life history stage of the crop during periods of drought. To better understand the ways in which crops respond to drought stress, we conducted a water limitation experiment with cultivated sunflower (*Helianthus annuus* L.). Six sunflower genotypes were subjected to moderate or severe drought stress over the vegetative or reproductive growth phases. Plants were phenotyped (growth rate, yield, water loss, and multiple leaf traits) and leaves were sampled for RNA-Seq. These data were used to investigate the extent to which phenotypic and gene expression responses to drought stress are specific to drought intensity and timing. Understanding how sunflowers cope with various stress scenarios should inform attempts to develop more drought-resilient cultivars.

PO0827: Oilseeds, Sunflower, and related

Transcriptomic Responses to Multiple Stressors in Wild and Cultivated Sunflower

Emily L. Dittmar and John M. Burke, University of Georgia, Athens, GA

Understanding plant responses to environmental stress may help inform breeding strategies aimed at producing more resilient crops. However, many studies that investigate the effects of stress are performed in controlled environmental conditions that alter a single stressor. Under natural conditions, plants may be exposed to a combination of different stressors acting simultaneously. Determining the extent to which plants exhibit novel responses when exposed to multiple stressors is a crucial step towards understanding how results from lab studies can inform plant breeding strategies aimed at producing robust plants under a variety of realistic field conditions. Of particular interest is understanding patterns of gene expression. Recent evidence suggests that the presence of multiple stressors operating simultaneously may produce novel patterns of gene expression that are not observed under individual stressors. The current study aims to understand whether two abiotic stressors (high salt and low nutrients) have interacting effects on plant performance and gene expression in the cultivated sunflower (*Helianthus annuus* L.).

The stress response of cultivated sunflower will be compared to that of its wild ancestor, (also *H. annuus*), and a wild relative adapted to coastal environments, *H. argophyllus*. Because it is believed that stress tolerance has been lost during domestication, efforts to breed more stress-tolerant crops often cross agricultural species with their wild relatives. However, the degree to which wild species have greater stress tolerance than domesticated species is often difficult to assess due to differences in growth form that preclude direct comparisons of fitness. Instead, this study will investigate gene expression in a comparative framework. Patterns of gene expression that have been conserved over evolutionary time-scales are likely to have beneficial effects on fitness. Identifying whether otherwise-conserved stress responses in wild species are not present in their domesticated relatives would provide evidence that stress tolerance has been lost during domestication and would identify candidate genes for plant breeding aimed at increasing stress tolerance in crops.

Finally, like many crops, domesticated sunflower exhibits heterosis which has led to its development as a hybrid crop. The current study will assess gene expression and performance in three sets of parental lines and their hybrid offspring to determine the degree to which heterosis may be explained by alterations in gene expression.

PE0828: Oilseeds, Sunflower, and related

Transcriptomics Analysis of Developing Seed from Wild Sunflower Populations across a Latitudinal Gradient

Max Barnhart, University of Georgia, Athens, GA

Sunflowers (*Helianthus annuus*) are native to North America and wild populations are distributed across a wide latitudinal gradient. Southern populations encounter different stresses than northern populations, including increased

heat and drought stress. As *H. annuus* is an important oilseed crop, we were particularly interested in how seeds differ across this latitudinal gradient. Our previous work found that seeds from southern populations had increased proportions of saturated fatty acids compared to seeds from northern populations. We grew seeds from Texan and Canadian wild sunflower populations in a common garden, then conducted a transcriptomics analysis of the developing seeds. Genes controlling the production and breakdown of palmitic acid were upregulated in developing seeds from Texas compared to developing seeds from Canada. This is concordant with the increased saturated fatty acid phenotype. Metabolic pathways for the production of pectin and response to bacterial pathogens were also upregulated in the developing seed from Texas sunflowers.

PO0829: Oilseeds, Sunflower, and related

Phylogeographic Insights into Sunflower Domestication

Brian Park and John M. Burke, University of Georgia, Athens, GA

Sunflower (*Helianthus annuus*) is one of the most important oilseed crops in the world. However, fairly little work has been done to reconstruct its recent evolutionary history, making it unclear where and when sunflower was domesticated. To this end, we used GBS to genotype a large panel of sunflower accessions, and used the resulting SNP data to infer evolutionary relationships among cultivated and wild populations and their respective demographic histories. We found that all cultivated sunflowers form a single, well-supported clade which is sister to wild populations from the Great Plains region (e.g., Nebraska and Kansas). Furthermore, we found that the cultivated sunflower lineage underwent a significant population bottleneck following their divergence from wild sunflower ~5,000-8000 years ago. Our results corroborate those from previous studies suggesting a single origin of cultivated sunflower in the central United States, and provide molecular insights into the timing of domestication.

PE0830: Oilseeds, Sunflower, and related

***Stevia rebaudiana*: Cultivated and Landraces Diversity Developing SSR and SNP Identification Through Reduced-Representation Library (RRL)**

Zoé Le Bihan¹, Galadriel Brière², Patrick Cosson³, Erwan Guichoux⁴, Cecile Hastoy⁵, Alexis Groppi², Luis Ernesto Errazu⁶, Dominique Rolin³, Raluca Uricaru⁷ and Valerie Schurdi-Levraud⁸, (1)INRA Bordeaux University, Biology and Fruit Pathology / Oviatis, Villenave d'Ornon, France, (2)Center of bioinformatic of Bordeaux (CBiB), Bordeaux, France, (3)INRA, Villenave d'Ornon cedex, France, (4)BIOGECO / Genomic center, Cestas, France, (5)Oviatis INRA, Lacauassade, France, (6)INTA-EEA, Famaillà, Argentina, (7)Informatic Laboratory of research in Bordeaux (LABRI), Talence, France, (8)INRA Bordeaux University, Villenave d'Ornon, France

Worldwide *Stevia rebaudiana* breeding improvement and genotype traceability suffer from a huge lack of information. Genomic information is particularly missing as no reference genome is available yet. Very few traceable and fixed cultivars are currently available for European producers and all have been selected outside of Europe. Therefore, *Stevia* modern breeding is a real challenge to offer adapted and elite cultivars to producers. This global objective means to be able to (1) classify germplasm in order to get knowledge on available genotypic and phenotypic variability and get tools to classify new coming germplasm (2) link genotypic and phenotypic variability in order to detect loci involved in main agronomic traits, such as biomass, SG content, favorable UGT alleles and response to pathogens, (3) understand the genetic architecture of the traits and develop marker assisted selection for future breeding purposes.

We developed SSR markers and SNP through Reduced-Representation Library (RRL) sequencing approaches and adapted SNP calling pipeline. All these markers were used to analyze the genetic diversity of 145 worldwide cultivated and landraces genotypes. They allow us to classify and analyze *Stevia rebaudiana* genetic diversity. They will be further used to create a genetic map and detect QTL associated with phenotype variation in offsprings.

Key words: *Stevia rebaudiana*; germplasm, molecular markers, SSR, SNP, sequencing

Davey JW, Cezard T, Fuentes-Utrilla P, Eland C, Gharbi K, Blaxter ML. Special features of RAD Sequencing data: implications for genotyping. *Mol Ecol. England*; 2013;22: 3151–3164. doi:10.1111/mec.12084

Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, et al. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One. United States*; 2011;6: e19379. doi:10.1371/journal.pone.0019379

PO0831: Oilseeds, Sunflower, and related

Characterization of High Oleate Mutants Identified from Cultivated Peanut and Its Wild Relatives

Ming Li Wang¹, Brandon Tonniss¹, Charles Y. Chen² and Shyamalrau P. Tallury¹, (1)USDA-ARS, PGRCU, Griffin, GA, (2)Department of Crop, Soil and Environmental Sciences, Auburn University, Auburn, AL

Peanut is one of the important oilseed crops in the world. Since high oleate ($\geq 80\%$) peanuts have an extended-shelf life and good seed nutritional quality with high antioxidant activity, development of high oleate cultivars is an important breeding objective. For identification of useful genetic materials from the germplasm, the entire USDA cultivated peanut germplasm collection and some wild species (in total over 9,000 accessions) were screened for high oleate mutants using gas chromatography (GC). High oleate mutants were identified in three cultivated peanut accessions and one wild species accession. Point mutations from *FAD2* genes were identified by sequence analysis. These mutants were regrown in the field or greenhouse for morphological observation. All three cultivated high oleate mutants were from the subspecies of *hypogaea* (no flowers on the main stem). Two of them also had a high level of resistance to leaf spots and tomato spot wilt virus (TSWV). The wild high oleate mutant was from *Arachis veigae* (formerly *Arachis sylvestris*). This wild mutant also had a high level of very long-chain ($C \geq 22$) fatty acids (25.8%). These unique accessions will be useful genetic materials for seed quality improvement and disease resistance in peanut breeding programs.

PE0832: Oilseeds, Sunflower, and related

Genome Wide Association Studies for Seed Traits in *Camelina sativa*

Huang Li¹, Xiao Hu¹, Wilson Craine², Paul Grabowski³, John Lovell³, Kevin King¹, Indika Kahanda¹, Brendan Mumey¹, John Browse², Scot Hulbert², Jennifer Lachowiec¹, Jeremy Schmutz³ and **Chaofu Lu**¹, (1)Montana State University, Bozeman, MT, (2)Washington State University, Pullman, WA, (3)HudsonAlpha Institute for Biotechnology, Huntsville, AL

Dissecting the genetic basis of *Camelina* seed quality and yield-determining traits is crucial for its agronomic development. *Camelina* seed size, seed weight and seed composition including oil, fatty acids and protein contents, are all complex polygenic traits. In this study, we performed whole genome resequencing of a diversity panel consisting 222 *Camelina* accessions at 30X depth. Approximately 162,000 high-quality SNPs were chosen and integrated with phenotypic data from multiple field trials for Genome-Wide Association Studies (GWAS). Population structure and principle component analysis suggested four distinct subpopulations in our collection. The GWAS results revealed a distributed pattern of associated loci across chromosomes with varied significance on seed size and weight, reflecting the quantitative nature of these traits. Among these, several major-effect loci of seed size and weight coincided with the QTL identified in our previous linkage mapping study, lending higher resolution for gene discovery. In a parallel analysis on fatty acids, we identified the *FAD2* gene (Csa01g013220.1) within the locus significantly associated with linoleic acid concentration. We also revealed hotspots on chromosome 3 and 12 were related to oil and protein contents, which hold promise for disentangling the complex relationship among oil, protein and other seed components in *Camelina*.

PO0833: Oilseeds, Sunflower, and related

Long-Read RNA Sequencing Reveals Transcriptome Complexity in *Sesamum indicum* L

Libin Wei¹, Hongmei Miao¹, Haiyang Zhang¹ and PAG Authors' Group, (1)Henan Sesame Research Center, Henan Academy of Agricultural Sciences, China, Zhengzhou, China

As an important oilseed crop, sesame (*Sesamum indicum* L., $2n = 26$), is known as the ‘Queen of the oil seeds’ for its high oil content and quality. In this study, the full length transcriptome of 8 samples including root, stem, leaf, bud and seeds with various development stages were sequenced using the long-read single-molecule Oxford Nanopore MinION sequencing technology (ONT RNA-seq) at the first time. In total, 104.44 million clean reads (138.32 Gb) were obtained with an average of 1,324 bp per read. A total of 7,034 novel genes and 86,776 novel transcripts were obtained after comparing with the sesame reference genome. Additionally, we identified 77,646 novel open reading frames (ORFs), 1,921 long non-coding RNAs (lncRNAs), and 8,724 transcription factors (TFs). Moreover, a total of 33 fusion transcripts, 93,300 alternative polyadenylation (APA) and 50,249 alternative splicing (AS) events were detected. Overall, our results not only offer a comprehensive view of sesame transcriptome, but also reveal the transcriptome complexity in sesame.

PE0834: Oilseeds, Sunflower, and related

Genomic Predictions Improve the Performance of Clonal Cultivars in Oil Palm

Achille Nyouma, University of Yaoundé I, Yaoundé, Cameroon and **Florence JACOB**, **Joseph BELL**, **Indra SYAHPUTRA**, **Dadang AFFANDI**, **Benoit COCHARD**, **Tristan DURAND-GASSELIN**, **David CROS**

Prediction of clonal genetic values is among the difficulties of the genetic improvement of oil palm (*Elaeis guineensis* Jacq.) yield. Presently, clonal selection requires two stages of phenotypic selection (PS): preselection on the phenotypic values of one or two yield components having high heritability, and final selection on performances in clonal trials. The current study evaluated the efficiency of genomic selection (GS) for clonal selection on eight traits. The GS models were trained on 295 and 279 Deli × La Mé crosses for bunch production and quality components, respectively, and were validated on 42 Deli × La Mé ortets of known clonal value. Genotyping by sequencing led to a dense genome coverage with 15,054 single nucleotide polymorphisms (SNP). We assessed the effects of SNP dataset (SNP density and quality) and of two GS modelling approaches on prediction accuracy. The results showed prediction accuracies that ranged between 0.69 and -0.07 according to trait, SNP dataset and model. Modeling the parental origin of alleles gave the highest prediction accuracies for the traits used to define the two oil palm heterotic groups (bunch number and average bunch weight). The greatest GS prediction accuracies were beyond those of PS for six traits out of eight. Prediction accuracies from 0.69 to 0.37 for all traits can be achieved using GS or PS, depending on trait. This will enable preselecting ortet candidates on all traits before clonal trials, thus increasing the selection intensity and the genetic progress.

PO0835: Oilseeds, Sunflower, and related

Understanding the Genetic Basis of Natural Seed Size and Oil Content Variation in Pennycress

Zenith Tandukar, Katherine Frels, Ratan Chopra and James A Anderson, University of Minnesota, St. Paul, MN

Pennycress is a candidate for direct domestication to serve as a winter annual cash cover crop for the US Midwest. Growing pennycress on fallow croplands provide excellent ecosystem services like reduced soil erosion, reduced nitrate and nutrient leaching, increased water holding capacity and soil organic matter, and buffering against pest and disease pressure. Despite these benefits, cover crop acreage is woefully low due to the lack of financial gains for farmers. Pennycress is a great option for farmers because of its extreme winter hardiness (up to -30°C), high natural seed yields (1500 – 2000 kg/hectare), highly oil-rich seeds (26%-39% wt/wt%), high disease resistance, and protein for bioplastics and other end uses. The development of pennycress varieties with consistently high yields and high oil content will set it up for more widespread adoption among farmers who can maximize their profits with a second cash crop with the same amount of land, that does not compete directly with their major annual cash crop (for e.g. corn, soybeans, etc.). Understanding the genetic basis of natural variation in important traits like seed size and oil content, and integrating them in the breeding program is vital to make progress towards developing the first successful variety release. For this purpose, we are developing three different recombinant inbred line mapping populations ($N = \sim 180$ / RIL population), and an association mapping population ($N = \sim 300$) to initiate mapping efforts for quantitative trait loci (QTL) for the first time in pennycress. The main objectives are: i) To explore the natural variation of seed size and oil content related traits, ii) To map QTL and identify candidate genes for selected traits, and iii) To construct the first genetic linkage maps in pennycress. The genetic linkage maps will be based on genotyping-by-sequencing (GBS) derived markers obtained from two different F_2 populations. The association

mapping population has been genotyped using GBS, and the derived markers will be used to perform association mapping for agronomic, physiological, seed-related, and oil content traits in multi-year, multi-location trials in Minnesota.

PE0836: Rice

Rice Transcription Factor Binding Atlas

Taiji Kawakatsu, Institute of Agrobiological Sciences, NARO, Tsukuba, Japan

Gene expression regulation plays central roles in all biological processes, such as development and response to environmental signals. Recognition of *cis*-elements by transcription factors (TFs) fundamentally shapes the gene regulatory networks. Comprehensive collection of TFs and their binding sites is critical for understanding of gene regulatory network in an organism. Here, we present *in vitro* TF binding atlas in rice, using DNA affinity purification sequencing (DAP-seq). We transferred >1000 rice TF cDNAs from Gateway entry clones to pIX-Halo vector. DAP-seq identified *in vitro* binding sites of hundreds of TFs. We also identified distinct recognition motives between closely related TFs. Collectively, we provide the novel resource for deciphering gene regulatory network in rice.

PO0837: Rice

GWAS Identifies a Gene Controlling Rice Architecture

Kenji Yano¹, Gen Tamiya¹, Shinya Yoshida², Hidemi Kitano³ and Makoto Matsuoka³, (1)RIKEN, Tokyo, Japan, (2)Research Institute for Food and Agriculture, Ryukoku University, Shiga, Japan, (3)Bioscience and Biotechnology Center, Nagoya University, Nagoya, Japan

Rice architecture is an important agronomic trait determining crop production of rice, but it is difficult to study its molecular biology because of its complex traits. This study applied a new strategy combining PCA and GWAS, and identified a novel gene, *SPINDLY* (*OsSPY*). Further study found that *OsSPY* is a major factor controlling rice architecture by *O*-fucosylation of DELLA protein. We demonstrate the effectiveness of a new strategy combining PCA and GWAS for studying complex traits.

PE0838: Rice

Genome Wide Analysis of a Fast Neutron Induced Rice Mutant Population and Its Reference Kitaake Genome

Rashmi Jain¹, Mawsheng Chern¹, Guotian Li², Jerry Jenkins³, Shengqiang Shu⁴, David Kudrna⁵, Joel Martin⁴, Anna Lipzen⁴, Kerrie W. Barry⁶, Rod Wing⁷, Jeremy Schmutz³ and Pamela Ronald⁸, (1)University of California, Davis, Davis, CA, (2)University of California, Davis, Davis, CA, (3)HudsonAlpha Institute for Biotechnology, Huntsville, AL, (4)DOE Joint Genome Institute, Walnut Creek, CA, (5)Arizona Genomics Institute, Tucson, AZ, (6)DOE Joint Genome Institute, Berkeley, CA, (7)Arizona Genomics Institute, University of Arizona, Tucson, AZ, (8)Joint BioEnergy Institute (JBEI), Emeryville, CA

The Kitaake variety (*ssp. japonica*), has emerged as a model for rice research. It is an extremely early flowering rice cultivar, easy to propagate, with good yield potential and eating quality. Here, we report the de novo genome sequencing and analysis of KitaakeX, a Kitaake variety carrying the XA21 immune receptor. Our KitaakeX sequence assembly contains 377.6 Mb, consisting of 33 scaffolds (476 contigs) with a contig N50 of 1.4 Mb. The assembly is complemented with detailed gene annotations of 35,594 protein coding genes. We identified 331,335 variations between KitaakeX and the reference genome Nipponbare (*ssp. japonica*), and 2,785,991 variations between KitaakeX and Zhenshan97 (*ssp. indica*). We also compared Kitaake resequencing reads to the KitaakeX assembly and identified 219 small variations. We have generated a Kitaake mutant database called KitBase (<http://kitbase.ucdavis.edu/>), which includes genomic data, phenotypic data, and seed information for the 2400 fast neutron-induced KitaakeX mutant lines. The availability of the high quality KitaakeX genome and annotation reported here will greatly facilitate analysis of the FN mutants which was previously compared to Nipponbare as the reference genome. We aim to finish 4000 mutant lines by the end of year 2020 which will facilitate rice research.

PO0839: Rice

Impact of Panicle Structure on Grain Weight Distribution in Japanese Rice Cultivars

Shiori Yabe¹, Hiroe Yoshida², Hiromi Kajiya-Kanegae³, Eiji Yamamoto⁴, Masanori Yamasaki⁵, Hiroyoshi Iwata⁶, Kaworu Ebana⁷, Erina Fushimi², Hideo Maeda¹, Takeshi Hayashi¹ and Hiroshi Nakagawa², (1)Institute of Crop Science, NARO, Tsukuba, Japan, (2)Institute for Agro-Environmental Sciences, NARO, Ibaraki, Japan, (3)Research Center for Agricultural Information Technology, NARO, Tsukuba, Japan, (4)Kazusa DNA Research Institute, Kisarazu, Japan, (5)Food Resources Education and Research Center, Graduate School of Agricultural Science, Kobe University, Kasai, Hyogo, Japan, (6)The University of Tokyo, Tokyo, Japan, (7)Genetic Resources Center, NARO, Tsukuba, Japan

Panicle structure may have an impact on grain-filling efficiency in rice via the sink capacities and transportation within a panicle. The relationship between panicle size (i.e., panicle length and the number of spikelets and branches in a panicle) and grain yield has been verified so far, and a number of QTLs controlling spikelet number have been detected. To improve the grain yield, however, the other factors of panicle structure such as branching pattern should be taken into consideration. In the present study, we estimated the relationship between panicle structure and grain weight distribution in 143 Japanese rice cultivars based on Bayesian multivariate Gaussian model and structural equation model using whole-genome marker data. The variation of panicle structure and grain weight distribution was investigated in six environments in Japan, and the relationship between the traits was estimated in each environment. For both panicle structure relating traits and grain weight distribution relating traits, the magnitude of change in phenotypic value caused by environmental variation varied among cultivars. In most environments, effect from panicle length to panicle branching pattern was detected. Both of panicle size and branching pattern could affect the grain weight distribution directly and indirectly. The estimated relationship network changed among the evaluated environments, and sign of each effect changed occasionally even when the same direction of effects was estimated. This result suggests that the phenotype and relationship between traits showed genotype by environment interaction.

PE0840: Rice

Profiling Na⁺ and K⁺ Accumulation Across Biparental Inbreds of Rice By Hyperspectral Imaging

Isaiah Pabuayon¹, Irish Lorraine Pabuayon², Harkamal Walia³, Glen Ritchie¹, Rakesh Kumar Singh⁴ and Benildo G. de los Reyes¹, (1)Department of Plant and Soil Science, Texas Tech University, Lubbock, TX, (2)Texas Tech University, Lubbock, TX, (3)University of Nebraska Lincoln, Lincoln, NE, (4)International Rice Research Institute, Los Banos, Philippines

Genotypes that exclude Na⁺ to maintain ion homeostasis are typically selected for salt tolerance potential in pre-breeding programs. Determination of ion content is typically done destructively using flame photometry on tissue samples. The nuances of how salt accumulates at the whole-plant and organ levels are lost, as the ion accumulation dynamics prior to the collection are unknown. To address this important limitation, a non-destructive phenotyping approach was explored to examine organ-level ion accumulation in real-time and to lessen the effort needed for data collection. We used hyperspectral imaging to predict the trends of Na⁺ and K⁺ accumulation in six (6) F₈-recombinant inbred lines (RILs) derived from IR29×Pokkali, displaying a gradient of salt tolerance. Images at wavelengths from 550 nm (green) to 1700 nm (shortwave infrared, SWIR) were taken from plants under saline (EC 9; ~90 mM NaCl) and control conditions for 18 days. A model was constructed with partial least squares regression (PLSR) using the spectral data at the 18th day and ion concentrations determined through flame photometry. An R² of 0.72 was achieved, and this was used to predict plant ion concentrations through the experiment. Higher accumulation was detected in the salt-sensitive RIL, while tolerant RIL have salt concentrations closer to their respective controls. Among the genotypes, the transgressive mega-tolerant RIL had the closest trend between stress and control and the transgressive mega-sensitive RIL had the farthest. Hyperspectral imaging as a potentially viable tool for non-destructive high-throughput phenotyping for ion accumulation *in planta*.

PO0841: Rice

Improving Fe/Zn Concentration in Japonica and Indica Rice By Barley Sucrose Transporter Overexpression

Thanh Hai Tran¹, My-my Huynh², Carly Schramm¹, James Stangoulis¹, Colin Jenkins¹ and Peter Anderson¹, (1)Flinders University, Adelaide, SA, Australia, (2)Phoenix Molecular Designs, Vancouver, BC, Canada

Biofortification of crops is a way to address malnutrition-induced micronutrient deficiency. Rice is a good candidate for micronutrient biofortification as it is a staple food for over 50% of the global population, and yet the rice endosperm has very low levels of micronutrients. Fe/Zn concentration has been increased in wheat grains by increasing sink strength via the endosperm-specified overexpression of a barley sucrose transporter, *HvSUT1*. A similar approach in Japonica (Nipponbare) and Indica (IR64) rice was hypothesised to improve Fe/Zn content in the rice grain. We show that in both rice cultivars, the barley sucrose transporter was expressed in the grain during critical time-points of nutrient uptake from 5 days after anthesis (DAA) to 15 DAA, and reached its highest expression at approximately 10 DAA. In transgenic Japonica rice grains, Fe and Zn were distributed deeper into the endosperm, compared to the non-transgenic grains. We now aim to measure the level and localisation of Fe and Zn in the grain of the transgenic Indica rice.

PE0842: Rice

Identification of Novel Mutations in Genes Involved in Metalloid Uptake and Accumulation in Rice (*Oryza sativa* L.)

Thomas H. Tai, USDA, Agricultural Research Service, Davis, CA and HyunJung Kim, Dept. of Plant Sciences, UC Davis, Davis, CA

Novel mutations in rice genes involved in silicon (Si) and arsenic (As) transport (*Lsi1*, *Lsi2*) and vacuolar sequestration of As (*OsABCC1*) were identified using TILLING by sequencing. A population of chemically-induced mutants (n = 2048) was screened resulting in the detection of 61 putative mutations. Following removal of mutations predicted to be synonymous or residing in introns, Sanger sequencing confirmed 21 nonsynonymous mutations and 13 M3 lines harboring homozygous mutations (three *Lsi1*, nine *Lsi2*, and one *Osabcc1*) were identified for phenotypic evaluation. Altered sensitivity to germanium (Ge), a phytotoxic analog of Si, was observed in three lines. NM-E1746 and NM-3403 (both *Lsi1*) had increased tolerance whereas NM-3036 (*Lsi2*) was more sensitive, however, this appears unrelated to the mutation. Analysis of the straw from field grown plants revealed that NM-E1746 and NM-3403 were the only lines with significant reductions in total Si. Both mutants also had significant increases in total As and NM-3403 exhibited higher grain total As. The third *Lsi1* mutant (NM-3380) and two *Lsi2* mutants (NM-2902 and NM-2249) had increased straw total As. Increased grain total As was observed in NM-2902, NM-2249, and a third *Lsi2* mutant NM-E2244. Interestingly, NM-4903 (*Osabcc1*) had the highest total Si and was also the only line to have significantly less straw and grain total As. Confirmation of these results from a second year of field grown plants is underway. These novel mutant alleles represent useful genetic resources for further dissection of Si, As, and Ge transport in rice and the corresponding germplasm has potential for enhancing rice productivity and quality.

PO0843: Rice

Genetic Diversity of Korean Landraces and Modern Varieties in Rice (*Oryza sativa* L.)

Yoon Kyung Lee, Phyu Phyu Maung and Hee-Jong Koh, Seoul National University, Seoul, South Korea

Crop breeding has faced bottlenecks, thus attempt to look for sources of genetic variation is crucial in achieving breeding goals. While modern temperate japonica rice has very low genetic diversity, landraces are highly diverse and possess valuable traits. Analysis on genetic diversity gives us clues to practically utilize such resources. For the analysis, 211 accessions of Korean rice were used in total - 134 improved varieties and 87 landraces. To unravel the genome of accessions, Illumina platform with ten fold-coverage was used in sequencing. Informative SNPs were selected from raw SNPs for PCA, structure analysis, and phylogenetic analysis. The accessions were mainly classified into two groups, indica/Tongil type group and japonica group. It was noticed that japonica accessions were further divided into subgroups of improved varieties and landraces. The identified genetic structure of Korean accessions and its further study on some specific genetic regions could provide useful genome information for improvement of temperate japonica rice.

PE0844: Rice

Characterization of QTL for Grain Quality and Agronomic Traits Using Recombinant Inbred Lines Derived from a japonica/japonica Cross in Rice

Cheryl Adeva, Chungnam National University, Daejeon, South Korea, Jun Hyun Cho, Department of Southern Area Crop Science, National Institute of Crop Science, Miryang, South Korea, Hyun-Sook Lee, Dept. of Agronomy, Chungnam National University, Daejeon, Korea, Republic of (South), Changsoo Kim, Department of Crop Science, Chungnam National University, Daejeon, Korea, Republic of (South) and **Sang-Nag Ahn**, Dept. of Agronomy, Chungnam National University, Daejeon, South Korea

In this study, we combined two genomic technologies, Kompetitive Allele-Specific PCR (KASP) approach and Genotyping-By-Sequencing (GBS) technology to develop a saturated map of rice and identify useful or novel genes present in japonica/japonica population. For this, we developed 92 (F₆ and F₇) recombinant inbred lines (RILs) derived from a cross between two temperate japonica cultivars, Dodam (high in resistant starch (RS)) and Hwayeong (a non-waxy cultivar) and evaluated two starch-related traits, four mineral elements, and agronomic traits. A total of 215 high-quality SNPs was used to construct a physical map. However, no SNP in some specific regions were detected resulting in big gaps of the RILs. Thus, 41 KASP markers were used to fill and analyze the gaps. QTL analysis identified QTLs for contents for RS, AC and mineral elements including Fe and heading date. These results demonstrate that GBS-KASP represents a combined approach to identify QTLs between closely-related japonica rice cultivars. The markers would be informative to develop rice varieties with increased nutritional value in rice breeding programs.

Acknowledgment: This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (2017R1A2B2007554).

PO0845: Rice

Characterization of Cuticle Wax-Deficient Mutants in Rice (*Oryza sativa* L.)

Timothy Butterfield, USDA-ARS-CPGR, DAVIS, CA, HyunJung Kim, Dept. of Plant Sciences, UC Davis, Davis, CA and Thomas H. Tai, USDA-ARS, Davis, CA

Epicuticular waxes form the outermost protective barrier of the aerial surfaces of land plants and work in concert with other components of the plant cuticle to prevent uncontrolled loss of water and provide protection against an array of external environmental stresses. Approximately 4,750 M2 families, derived from chemical mutagenesis of rice seeds (cv. Sabine) were screened for adhesion of water droplets to leaves resulting in a wet leaf/glossy (*w/g*) phenotype. Sixteen independently derived mutants with altered water adhesion were identified. All the mutants except SAB-1558 exhibited the *w/g* phenotype due to coalescence of water droplets. The SAB-1558 was characterized by the adherence of small, discrete water droplets. An additional *w/g* mutant was identified in the mutant line KDS-2249D, which is derived from the cv. Kitaake. Preliminary genetic analysis revealed that at least five of the 17 mutants result from single gene recessive mutations. Targeted exon capture and sequencing identified a nonsense mutation in the *OsGLI-1* gene in KDS-2249D. SEM analysis has confirmed the association of the *w/g* phenotype with a deficiency in epicuticular wax crystals. In contrast, the SAB-1558 mutant appears to have larger epicuticular wax crystals but at much lower density than wild type. Progress towards identification of the underlying mutations, wax composition and quantification using GC/MS analysis, and phenotypic evaluation of the effects of the cuticular wax deficiencies of these mutants on stress tolerance will be presented.

PE0846: Rice

A Genomic Pipeline for Developing Genome Edited Rice for Shortened Days to Flowering

Karina Y. Morales¹, Benjamin Rogers¹, Haley Sloan¹, Blake Pope¹, Emily Chao¹, Rodante E. Tabien² and Michael J. Thomson¹, (1)Texas A&M University, College Station, TX, (2)Texas A&M AgriLife Research, Beaumont, TX

Rice is an essential food crop providing the daily caloric intake of over 50% of the world's population. Flowering is one of the most sensitive stages of rice growth and is highly variable among varieties and across environments. In Texas, early maturing varieties are desired as they can avoid flowering in the hottest part of the summer and allow the ratoon crop to mature before cold temperatures set in. In order to develop shorter flowering rice varieties, a diversity panel was genotyped using both the 7K Cornell-IR LD Rice Array and AgSeq, a skim sequencing method. A genome wide association study (GWAS) was performed using phenotypic data from the diversity panel grown in Beaumont, TX in 2018 and 2019. Candidate genes identified from our GWAS will be validated using CRISPR/Cas genome editing in an elite variety from Texas in a multiplexed fashion to identify how they act individually and in combination with each other. Finally, a custom multiplex amplicon sequencing panel is being developed for characterizing diversity of flowering time candidate genes and validation of target genes in the genome edited progeny.

PO0847: Rice

Genetic Structure Analysis of 48 Modern Varieties in Rice (*Oryza sativa* L.)

Yijie Gui¹, Weiqi Tang², Shaohua Yang¹, Zaijie Chen¹, Jianmin Chen¹, Weiren Wu³ and Feng Wang¹, (1)Institute of Biotechnology, Fujian Academy of Agricultural Sciences, Fuzhou, China, (2)Marine and Agricultural Biotechnology Laboratory, Institute of Oceanography, Minjiang University, Fuzhou, China, (3)Key Laboratory of Crop Genetics, Breeding and Multiple Utilization, Ministry of Education, Fujian Agriculture and Forestry University, Fuzhou, China

Modern rice varieties harbored elite breeding signature and superior alleles. Decoding the genetic structure of breeding accessions could provide valuable information for rice genetic improvement. In this study, 48 rice varieties including 31 *Indica* cultivars and 17 *Japonica* cultivars were selected for resequencing. Around 50G rice genomic data were produced using Illumina HiSeq X Ten platform, which covered on average 25-fold genome for each sample. Totally, 3,303,122 informative SNPs and InDels were generated, which were combined with the 3K rice genome core SNP data set to analyze the PCA, genetic structure and phylogenetic tree of these modern cultivars. The results showed that among the 48 accessions, 31 *Indica* varieties belonged to XI-1B group and the rest 17 *Japonica* varieties belonged to temperate *Japonica*.

PE0848: Rice

Synthetic Apomixis in Hybrid Rice

Kejian Wang, China National Rice Research Institute, Hangzhou, China

Heterosis, or hybrid vigor, is exploited by breeders to produce elite high-yielding crop lines but beneficial phenotypes are lost in subsequent generations owing to genetic segregation. Apomixis, or clonal propagation through seeds, would enable self-propagation of F₁ hybrids and permanently fix heterosis. Here, we established synthetic apomixis in hybrid rice using genome editing. We fixed the heterozygosity of F₁ hybrid rice by multiplex CRISPR-Cas9 genome editing of the *REC8*, *PAIR1* and *OSD1* meiotic genes to produce clonal diploid gametes and tetraploid seeds. Next, we showed that editing the *MATRILINEAL* (*MTL*) gene could induce formation of haploid seeds in hybrid rice. Finally, we combined fixation of heterozygosity and haploid induction by simultaneous editing of all four genes (*REC8*, *PAIR1*, *OSD1* and *MTL*) in hybrid rice, and obtained plants that could propagate clonally through seeds. Application of synthetic apomixis strategy may enable self-propagation of a broad range of elite F₁ hybrid crops.

PO0849: Rice

Using a Comparative Transcriptomics Approach to Identify Common Gene Expression Patterns in Rice Roots during Associations with Plant Growth-Promoting Bacteria

Jacklyn Thomas, Grant Wiggins, Connor Deen and **Arijit Mukherjee**, University of Central Arkansas, Conway, AR

Non-legume plants such as rice and maize can form beneficial associations with plant growth-promoting bacteria (PGPB) such as *Herbaspirillum seropedicae* and *Azospirillum brasilense*. Several studies have shown that these PGPB promote plant growth via nitrogen-fixation and hormone synthesis. Our current understanding of the molecular aspects and signaling that occur between plants like rice and PGPB like *Herbaspirillum seropedicae* is limited. In this study, we used an experimental system where the *H. seropedicae* could colonize the plant roots and promote plant growth in wild type rice. We then used this colonization model to identify the regulation of gene expression in rice roots at 1 day post-inoculation (dpi) with *H. seropedicae*. We identified 1688 differentially expressed genes (DEGs) in rice roots. We used singular enrichment analysis (SEA) with agriGO and identified 30 GO terms that were significantly enriched. These included 18 in biological processes (e.g., response to stimulus, signaling, etc.), 7 in molecular functions (e.g., catalytic activity, kinase activity, etc.), 5 in the cellular component (e.g., extracellular region, cell wall, etc.). We performed a comprehensive data mining to classify the DEGs into the categories of transcription factors (TFs), protein kinases (PKs), and transporters (TRs). Several of these DEGs encode proteins that are involved in the flavonoid biosynthetic pathway, defense, hormone signaling pathways, and nitrate and sugar transport. Next, we compared the DEGs identified in this study to those identified in rice roots during associations with *Azospirillum brasilense*, at 1dpi. We identified 148 DEGs that were upregulated in expression in both datasets. The expression pattern of these genes suggests that some of these are likely to play a role during associations with both *H. seropedicae* and *A. brasilense* and are excellent targets for future genetic studies.

PE0850: Rice

Climate-Ready Rice for South America: The Search for Stress Response Genes

Antonio Costa De Oliveira, Universidade Federal de Pelotas, Pelotas-RS, Brazil

Climate changes and increasing biotic and abiotic stresses are affecting crops worldwide. Rice, one of the top three cereals, is under constant constraint by these threats. Brazil, a very important rice producer and the largest producer outside Asia, has an annual production of over 10 million tons. Many stresses affect yields and lower total production each year. Stresses such as cold, flooding, drought, iron toxicity and salinity are constantly affecting farmer's yields. Our lab has been working in developing stress resilient lines to these different stresses. The understanding of plant response mechanisms is key to the development of stress resilient crops. WRKY transcription factors (TFs) are responsible for the regulation of genes responsive to many plant growth and developmental cues, and are involved in biotic and abiotic stress responses. Recently, functional genomics studies in model plants have enabled the identification of function and mechanism of action of several WRKY TFs in plants. Our group has been using molecular tools and mutation breeding to accelerate breeding for abiotic stress tolerance.

PO0851: Rice

Genetic Male Sterility Facilitated Recurrent Selection for Developing Multiple Stress Resistance Restorer Lines in Rice

Revathi Ponnuswamy, Crop Improvement-Hybrid rice section ICAR - Indian Institute of Rice Research (IIRR), Hyderabad, India

The use of recurrent selection in self pollinated crops is limited due to difficulty in making random crossing for recombination in each recurrent selection cycle. The monogenic recessive male sterile gene *ms* is very useful in facilitating inter crossing phase of rice recurrent selection cycle. In IR36 genetic male sterile (GMS) restorer population sterile plants were selected and crossed with donors of *Xa21* for BLB resistance, *Pikh* for blast resistance, *Bph20* & *21*, *Gm4*, *Gm8* for BPH and gall midge resistance. The F₁S plants were confirmed for the presence of resistance genes with the help of molecular markers. The bulked F₂ seeds were raised in isolation for random mating. Seeds from only sterile plants (pollinated by fertile plants) were harvested for further advancement. In each cycle MAS for the presence of *Xa21*, *Pikh*, *Bph18*, *Bph21*, *Gm4* and *Gm8* was done. The positive fertile plants were advanced by pedigree method from 5th cycle. Lines with five resistance genes *Pikh*, *Bph18* & *21*, *GM4* and *GM8*, four, three and single resistance gene were identified and their fertility restoration was tested by crossing with CMS lines to identify multiple stress resistance restorers.

PE0852: Rice

Rice Seedling Cold Stress QTL Revealed By Genome-Wide Association Mapping and Biparental Mapping

Georgia C Eizenga¹, Naoki S. Shimoyama², Aaron K. Jackson¹, Jeremy D. Edwards¹, Avery R. Jessel² and Michael R. Schläppi², (1)USDA-ARS Dale Bumpers National Rice Research Center, Stuttgart, AR, (2)Marquette University, Milwaukee, WI

Improved seedling cold stress tolerance would allow rice to be planted earlier in the growing season in the USA, thus increasing yield. To identify genes controlling seedling cold tolerance, the USDA Rice Minicore (RMC) and Rice Diversity Panel 1 (RDP1) association mapping panels were phenotyped for cold stress traits designed to mimic the natural environment, including low temperature seedling survivability (LTSS) and electrolyte leakage (EL). Genome-wide association (GWA) mapping was conducted with genotypes derived from 3,200K SNPs for the RMC and 700K SNPs for the RDP1. For RDP1, 40 GWA-QTL were discovered and six Multiple Chilling Phenotype (MP)-QTL, *qMP3-1*, *qMP6-2*, *qMP9-4*, *qMP10-1*, *qMP10-4* and *qMP11-2* overlapped with previously reported RMC GWA-QTL. Database searches revealed eleven candidate genes near *MP3-1*, *MP6-2*, *MP9-4* and *MP10-4*. To further validate these GWA-QTL, both recombinant (RIL) and backcross (BIL) inbred line mapping populations were developed from three RMC accessions, Krasnodarskij 3352, Wir911 and Carolino 164. The two populations were genotyped with 7K SNPs and phenotyped for LTSS and EL. For the RILs, six LTSS-QTL and one EL-QTL were found on chromosomes 3, 4, 6, 8 and 9. Only *qLTSS6* did not have an overlapping GWA-QTL. Four putative candidate genes associated with various stress tolerances, were identified in four of the QTL regions. Only *qEL6* was specifically associated with cold stress. For the BILs, three possible QTL, *qLTSS1*, *qLTSS8* and *qEL8* were found, but only *qLTSS8* was near a GWA-QTL. The next steps will be to validate selected candidate genes and develop markers for breeding.

PO0853: Rice

Rice Decussate Mediates Yield Retention Under Drought and Provides Convincing Evidence of Functional Interactions Between *qDTY12.1* and the Resident Genomic Background

Jacobo Sanchez¹, Pushpinder Pal Kaur¹, Naga Bhushana Rao Karampudi¹, Ai Kitazumi¹, Isaiah Pabuayan¹, Nitika Sandhu², Margaret Catolos², Arvind Kumar² and Benildo G. de los Reyes¹, (1)Department of Plant and Soil Science, Texas Tech University, Lubbock, TX, (2)International Rice Research Institute, Los Banos, Philippines

Breeding efforts in rice to improve yield retention under drought have been effective with the discovery and introgression of large-effect QTL (*qDTY*), specifically, *qDTY12.1*. Previous efforts to identify the alleles that drive the phenotype have mainly concentrated on intra-QTL interactions. The likelihood of interactions between *qDTY12.1* allele(s) and the recipient parent's genetic background spurred our analysis of the transcriptomes of two phenotypically distinct sibling backcrossed introgression lines (BILs), the original *qDTY12.1* donor (Way Rarem) and cv. IR64. Results showed growth-stage dependent transcriptomic signature differences and similarities between the superior BIL and the other three genotypes. Specifically, the superior BIL had a unique transcriptomic signature during the booting stage of development under both drought and well-irrigated conditions. Booting-stage expression analysis revealed that an allele within *qDTY12.1*, namely Decussate, anchors a genetic circuit relevant to yield retention and whose member alleles reside outside the *qDTY12.1* region. Allelic vestiges from Way Rarem as well as from IR64 were visible in the transcriptomes of both BILs showing that the genetic circuitry post-introgression was still heterogeneous despite several rounds of backcrosses to IR64. This genetic heterogeneity in the superior BIL along with the Decussate allele within *qDTY12.1*, we believe, provides for the unique genetic combination and network circuitry that gives rise to yield retention under reproductive stage drought.

PE0854: Rice

Molecular Genetic Analysis of Drought Resistance and Productivity Mechanisms in Rice

Yheni Dwiningsih, Julie Thomas, Anuj Kumar, Chirag Gupta, Charles Ruiz, Sara Yingling, Elizabeth Crowley and Andy Pereira, Department of Crop, Soil and Environmental Sciences, University of Arkansas, Fayetteville, AR

Rice is a major world food crop with USA being the third largest exporter of rice, and Arkansas accounting for more than 40% of U.S. rice production. Rice production in Arkansas is dependent on water reservoirs for stable irrigation. Rice productivity is dependent on water availability, with rice using 2-3 times more water than other food crops. The objectives of the research were: 1) screening and dissection of drought resistance mechanisms in diverse rice genotypes and a recombinant inbred line (RIL) population; 2) identify molecular markers for drought resistance (DR) traits in the segregating population. The rice genotypes exhibited differential drought resistance mechanisms categorized as drought tolerance, drought avoidance, and drought escape on the basis of morphological and physiological differences (ABA response). The documented morphological, physiological, and grain yield parameters differed significantly between the genotypes ($P \leq 0.01$) and treatments ($P \leq 0.01$) with a strong genotype x treatment interaction ($P \leq 0.01$). A RIL population derived from the varieties Kaybonnet (DR) and ZHE733 (sensitive), termed K/Z RILs, was chosen for molecular genetic analysis of drought resistance traits. The RIL population was screened using controlled drought stress treatment at the reproductive stage in the field, and the effect of stress quantified by the number of filled grains per panicle. Based on the DR scores, a genetic screen was done using bulked segregant analysis (BSA), where sets of 10 DR and sensitive RIL plants were used for screening of SSR markers to find polymorphisms linked to the yield-related traits under drought. From this BSA screen, 6 polymorphic markers were identified: RM9 (Chr 1), RM109 (Chr 2), RM236 (Chr 2), RM114 (Chr 3), RM131 (Chr 4) & RM139 (Chr 11). Our results provide valuable information for dissecting the genetic basis of drought resistance mechanisms and provide a valuable resource for breeding US rice cultivars for a water saving agricultural system.

PO0855: Rice

Regulatory Network Mediated By the Rice Decussate Gene (DEC) Under Drought Stress

Oluwatobi R. Bello, Texas Tech University, Lubbock, TX

Identification and characterization of regulatory genes are important for understanding the complex genetic complex network that defines phenotypic expression potential under a given environmental condition. The *Decussate* (*DEC*) gene was initially discovered as a regulator of leaf phyllotaxy. Our recent studies showed that one of the *DEC* genes in rice is part of *qDTY12.1*, a QTL with major effects in yield maintenance under drought. *DEC-qDTY12.1* expression is important for the maintenance of flowering time under drought. The parents (IR64, Way Rarem) and the backcross introgression lines (BILs) in IR64 displaying high (HPB) and low (LPB) yield penalty under drought were used in this study. BILs carrying the Way Rarem allele showed robust *DEC* expression at the booting stage when the flower organs develop. Expression was further enhanced under drought stress. However, LPB had much higher expression compared to its siblings. A putative interacting gene (*OsC3H56*) was found to be upregulated during the booting stage and under drought only in the LPB. Sequence analysis showed that the BILs and Way Rarem showed the same *DEC* allele, indicating proper introgression of *qDTY12.1*. However, *OsC3H56* sequences differ between the genotypes examined. Promoter analysis of the different *DEC* alleles showed unique stress-inducible cis-elements in the Way Rarem allele. Collectively, these results suggest that *DEC* expression is important for maintained yield under drought. However, to be fully functional, *DEC* needs its interacting partners, one of which is possibly *OsC3H56*. The authors will present progress in characterizing the *DEC* partners in the regulatory network.

PE0856: Rice

Marker-Assisted Backcrossing for Introgression of the Saltol Locus Conferring Salt Stress Tolerance in Rice

Caterina Marè¹, Elisa Zampieri², **Alessandro Tondelli**¹, Viviana Cavallaro³, Franca Finocchiaro¹, Gianni Tacconi¹, Stefano Monaco², Marco Canella², Massimo Gennaro², Brigitte Courtois⁴, Julien Frouin⁴, Patrick Samper⁴, Laurent Brottier⁴, Luigi Cattivelli¹ and Giampiero Valè⁵, (1)CREA - Research Centre for Genomics and Bioinformatics, Fiorenzuola d'Arda, Italy, (2)CREA - Research Centre for Cereal and Industrial Crops, Vercelli, Italy, (3)Università degli Studi di Milano, Dip. di Scienze Agrarie e Ambientali, Milano, Italy, (4)CIRAD - UMR AGAP, Montpellier, France, (5)Università del Piemonte Orientale, Dip. di Scienze e Innovazione Tecnologica, Vercelli, Italy

Soil salinization represents a threat for rice cultivation. The H2020 project NEURICE (New Commercial EUropean Rice) aimed at identifying and introducing genetic variation for salt tolerance in European rice germplasm, mainly by exploiting the positive effect of the *Saltol* QTL in maintaining the Na/K homeostasis. The *Saltol* QTL from the *indica* rice donor IR64-Saltol (located on chromosome 1) was introgressed into two *japonica* Italian varieties, Onice and Vialone Nano following a marker-assisted backcross (MABC) scheme, through three backcrosses and two selfing to achieve the BC3F4 generation. During the backcrosses, the scheme was coupled to an embryo rescue technique to fasten the process. At each backcross cycle, the *foreground* and *background* selections relied on SNP-based KASP markers. The BC3F1 selected lines showed a 91-98 and 93-98 recovery percentage for Onice and Vialone Nano backcrosses, respectively. BC3F2 lines were genotyped to identify homozygous lines at *Saltol* locus and the best BC3F3 lines (10 in Vialone Nano and 12 in Onice) were subjected to genotyping by sequencing (GBS) to allow a more precise screening of the recurrent parent genome. Finally, the best BC3F4 lines were subjected to field phenotyping in salinized fields on delta Po river, to *in vivo* assess their salt tolerance.

PO0857: Rice

Positive and Negative Complementation Effects Determine Transgressive Salt Stress Tolerance in Recombinant Inbreds of Rice

Isaiah Pabuayon¹, Ai Kitazumi¹, Harkamal Walia², Rakesh Kumar Singh³, Glenn B. Gregorio⁴ and Benildo G. de los Reyes¹, (1)Department of Plant and Soil Science, Texas Tech University, Lubbock, TX, (2)University of Nebraska Lincoln, Lincoln, NE, (3)International Rice Research Institute, Los Baños, Philippines, (4)International Rice Research Institute, Los Baños, Laguna, Philippines

The new plant breeding paradigm has a grand challenge of creating novel adaptive phenotypes that have not been achieved before, in order to contribute an innovative solution to food security in the 21st century. Methods for creating optimal varieties for different environmental conditions needs to be re-envisioned. In this study, extremely salt-tolerant and salt-sensitive recombinants of rice were identified from an F₈ recombinant inbred population derived from a cross between salt-sensitive *indica* (cv. IR29) and salt-tolerant *aus* (cv. Pokkali) parents. A minority of recombinants showed tolerance that are beyond the phenotypic ranges of the parents. The mega-sensitive recombinant FL499 was unique in comparison to the rest of its siblings based on a battery of physiological traits involved in maintaining ionic homeostasis. On the other hand, the mega-tolerant recombinant FL510 was more of a hybrid between the two parents. We inferred that complementation occurred between parental beneficial alleles, creating a stronger phenotype hence reduction of physiological drags. Hyperspectral image analysis showed minimal change between stress and control in FL510, thus providing a mechanistic basis for its acquired salt tolerance potential. This is reflected in the overall transcriptome configuration, which showed very gradual changes compared to the rest of the population. Metabolome profiles also showed significant similarity to IR29 profile, further indicating that despite its inherent sensitivity, IR29 does have important contributions to the expression of superior tolerance potential. These results indicate that genome shuffling led to various combinations of beneficial and detrimental alleles in transgressive individuals.

PE0858: Rice

QTL Mapping for Resistance to *Scirpophaga Incertulas* Using *indica* Rice Recombinant Inbred Lines

Chuan xu Li, HZAU(Huazhong Agricultural University), Wuhan, China

Abstract: *Scirpophaga incertulas* is one of the most destructive pest of rice and mainly drill and eat rice stems, resulting in reduced yield. There are few resistant species against the widespread *Chilo suppressalis* and *Scirpophaga incertulas* in germplasm resources. We found that the *indica* rice A232 is moderately resistant to *Chilo suppressalis* and *Scirpophaga incertulas*. Our research was to clone the resistance genes (QTLs) of A232 to the rice *Scirpophaga incertulas*. We constructed a recombinant inbred line(F12) using A232 and Gang46B (susceptible *indica* rice varieties to *Scirpophaga incertulas*). We evaluated the level of resistance to *Scirpophaga incertulas* for 265 recombinant inbred lines under natural rice fields condition in Hainan province by measuring the dead-heart index (DHI) in 2012 and 2013. Among A232 and Gang46B, a total of 833309 SNPs were detected, including 644062 SNP markers with a depth of at least 4X that were applicable to RIL. A High density linkage map from this

population was constructed by using HighMap(HighMap is a high-density genetic map construction method based on high-throughput sequencing independently developed by BMK). The SNP markers in the linkage map were 232600 and the Bin was 3327. In the 12 linkage groups, the total genetic map distance was 2136.78cm, and the average genetic map distance was 0.64cm. Two QTLs were identified by R/qtl, namely qSBR-2 and qSBR-4, which were located on chromosomes 2 and 4 respectively. Among these QTLs, qSBR-4, mapping in the interval between Bin25600 and Bin25602 on chromosome 4, it had a relative high logarithm of the odds (LOD) score and explained 20.12% of the phenotypic variation. In the candidate gene region of chromosome 4, we selected seven candidate genes for transgenic verification, and preliminarily confirmed the candidate genes for insect resistance.

PO0859: Tomato, Potato, Pepper, and related

Multi-Omics Approaches Reveals Co-Evolution of microRNAs and Disease-Resistant Genes in Solanum Plants

JeongYeon Yoon, Seoul National University, Seoul, South Korea, Chanseok Shin, Seoul National University, Seoul, Korea, Republic of (South) and Authors' Group

MicroRNAs play roles in various biological processes like defense in plants. Some plant microRNAs produce phased, secondary siRNAs (phasiRNAs), which regulate cascade of gene expression. We investigated the relationship of Solanaceae microRNAs and defense genes in evolutionary perspective by performing genome-wide comparative analysis of microRNAs and their targets in Solanum plants (pepper, tomato, and potato). Degradome analysis showed that many of genes related to defense response are regulated by microRNAs in Solanum plants. We found that novel pepper microRNAs targeted genes encoding nucleotide-binding leucine rich repeat (NLR) or receptor-like protein, known as disease-resistant genes. In addition, these novel microRNAs triggered phasiRNA production indicating amplification of regulation of the disease-resistant gene families. Among these, miR-n033a-3p, whose target NLRs have been duplicated in pepper, targets more NLRs belonging to specific subgroup in pepper than those in potato. Taken together, microRNAs targeting resistance genes might have evolved to regulate numerous targets in Solanaceae, following expansion of target resistance genes. This study provides an insight into evolutionary relationship between miRNAs and their target defense genes in plants.

This work was supported by the Next-Generation BioGreen 21 Program (No. PJ01333001), Rural Development Administration, Republic of Korea.

PE0860: Tomato, Potato, Pepper, and related

Structural Variant Landscapes in Tomato and Their Role in Evolution, Domestication, and Breeding

Michael Alonge¹, Xingang Wang², Matthias Benoit², Melissa Kramer², Srividya Ramakrishnan¹, Sebastian Soyk², Daniel Ciren², Lei Zhang³, Lara Pereira³, Nathan Taitano⁴, Jennifer Kim^{2,5}, Yuval Levy⁶, Tom Harel⁶, Ziva Amsellem⁶, T. Rhyker Ranallo-Benavidez⁷, Sara Goodwin², W. Richard McCombie², Joyce Van Eck⁸, Yuval Eshed⁹, Esther van der Knaap¹⁰, Fritz Sedlazeck¹¹, Michael Schatz^{1,2} and Zach Lippman^{2,5}, (1)Johns Hopkins University, Baltimore, MD, (2)Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, (3)Center for Applied Genetic Technologies, Athens, GA, (4)University of Georgia, Athens, GA, (5)Howard Hughes Medical Institute, Cold Spring Harbor, NY, (6)Department of Plant and Environmental Sciences, Rehovot, Israel, (7)Department of Biomedical Engineering, Baltimore, MD, (8)Boyce Thompson Institute for Plant Research, Ithaca, NY, (9)The Weizmann Institute Of Science, Rehovot, Israel, (10)Department of Horticulture, Institute of Plant Breeding, Genetics and Genomics, University of Georgia, Athens, GA, (11)Baylor College of Medicine, Houston, TX

Structural variants (SVs) underlie many important evolution, domestication and breeding traits. A comprehensive understanding of SVs is critical to realize advanced crop improvement with modern breeding and engineering techniques, such as CRISPR/Cas9 gene editing. Short-read sequencing has limited the study of natural variation to small variants, which has often implicated associated rather than causal variants in studies relating genotype to phenotype. We collected over 8Tbp of long-read Oxford Nanopore sequence data from 100 diverse tomato accessions including wild progenitor (*S. pimpinellifolium*), early domesticated (*S. lyc. var. cerasiforme*) and modern cultivated (*S. lycopersicum*) lines. Using a combination of read-mapping and *de novo* assembly, we established the most comprehensive database of SVs in tomato to date. These data reveal insights into repeat content and activity,

introgressions and subclade diversity, including many instances of SVs intersecting genes and/or *cis*-regulatory regions associated with gene expression changes. We highlight three examples of SVs that have impacted important traits: First, a deletion within a cluster of glycosyltransferase genes is likely the causal variant for variation in guaiaicol and methylsalicylate, two metabolites that confer an undesirable smoky aroma in fruits. Next, we show that a multi-gene duplication is a causal variant for a major domestication fruit weight QTL. Finally, we reveal a complex SV locus that interacts with two additional transcription factor genes to regulate inflorescence branching. Our analyses underscore the important role of SVs in tomato, and we expect the approaches we have developed will be broadly applicable across many plant and animal species.

PO0861: Tomato, Potato, Pepper, and related

Accuracy of Genomic Selection for Improving Fruit Traits in Tomato (*Solanum lycopersicum* L.)

Minkyung Kim and **Sung-Chur Sim**, Sejong University, Dept. of Bioresources Engineering, Seoul, South Korea

Genomic selection (GS) is a breeding strategy to increase genetic gain per cycle using genomic estimated breeding values (GEBVs). For prediction of GEBVs, several GS models were used in crop species. In this study, we compared GEBVs from three GS models including BayesA, Bayesian LASSO, and Ridge Regression for tomato traits. The phenotypic variations of eight fruit traits (weight, width, length, shape index, hardness, pericarp thickness, locule number, and Brix) were evaluated in a breeding population (n=162) with three replications over two years. This population was genotyped using the 55K Axiom™ SNP array. A total of 24,897 SNPs were pre-filtered using the missing rate (< 0.05) and minor allele frequency (> 0.1). Of these, 1,536 SNPs were selected based on their physical positions for further analysis. Prediction accuracy of the GS models were assessed based on correlation between GEBVs and phenotypes using the leave-one-out cross validation (LOOCV) and K-fold methods in R packages. The Pearson correlation coefficients of the 2018 phenotypes were higher than those of the 2019 phenotypes for all eight traits. However, similar levels of prediction accuracy were observed between the GS models in both LOOCV and K-fold methods. The three GS models showed relatively high correlation coefficients for pericarp thickness (0.619-0.860), fruit width (0.590-0.860), and fruit length (0.606-0.826) in the LOOCV method. In addition, lower correlation coefficients (0.362-0.460) were found for fruit hardness compared to the other traits. These results will be useful to increase the accuracy of genomic selection in tomato breeding programs.

PE0862: Tomato, Potato, Pepper, and related

Genomic Evidence for Complex Domestication History of the Cultivated Tomato in Latin America

Hamid Razifard, University of Massachusetts, Amherst, MA, Alexis Ramos, Auburn University, Auburn, AL, Denise Tieman, University of Florida, Gainesville, FL, Esther van der Knaap, Department of Horticulture, Institute of Plant Breeding, Genetics and Genomics, University of Georgia, Athens, GA and Ana Caicedo, Biology Department, University of Massachusetts, Amherst, MA

Plant domestication is often a complex process, involving under-explored intermediate stages with important implications for the evolutionary trajectories of domestication traits. Previously, tomato domestication history was thought to involve two major transitions: one from wild *Solanum pimpinellifolium* L. (SP) to a semi-domesticated intermediate, *S. lycopersicum* L. var. *cerasiforme* (SLC), in South America, and a second transition from SLC to fully domesticated *S. lycopersicum* L. var. *lycopersicum* (SLL) in Mesoamerica. Using a dataset of 296 accessions obtained from de-novo whole-genome sequencing and from previously published data, we employ population genomic methods to reconstruct tomato domestication history, focusing on the evolutionary changes occurring in the intermediate stages. Our results suggest that SLC originated approximately 80 thousand years ago (KYA) predating earliest archaeological records of plant domestication, suggesting that SLC may have diverged from SP as a wild population. Many traits considered typical of cultivated tomatoes arose in South American SLC, but were lost or diminished once these partially domesticated SLC spread northward to Mesoamerica, likely due to increased fertility or adaptation to new environments. Several traits were reselected in a convergent fashion in SLL, the common cultivated tomato, probably around 7 KYA, prior to its expansion around the world. Based on these findings, we reveal complexities in the intermediate stage of tomato domestication and provide insight on trajectories of important genes and phenotypes involved in the tomato domestication syndrome. Our results also allow us to identify underexplored germplasm that harbors useful alleles for crop improvement.

PO0863: Tomato, Potato, Pepper, and related

Genetic Control of Reproductive Traits in Tomato Under High Temperature

Maria José Gonzalo¹, Yi-Chen Lee², Kaiyi Chen², David Gil³, Teresa Montero³, Inmaculada Nájera⁴, Carlos Baixauli⁴, Antonio Granell⁵ and **Antonio J. Monforte**⁵, (1)IBMCP. CSIC-UPV, Valencia, Spain, (2)National Taiwan University, Taipei, Taiwan, (3)Enza Zaden España S.L., Sta. M^a del Aguila, Spain, (4)Centro de Experiencias de Cajamar, Paiporta, Spain, (5)IBMCP, Valencia, Spain

Global climate change is increasing the range of temperatures crop plants face during their life cycle and currently is already having negative effects in yield. In this scenario, understanding the genetic control of plant responses to a range of increasing temperature conditions is a prerequisite for efficiently developing cultivars with increased resilience. The current work reports the identification of Quantitative Trait Loci (QTL) involved in a number of tomato traits that are affected by temperature during plant growth. Reproductive traits such as number of flower (FLN) and fruits (FRN) per truss and percentage of fruit set (FRS), pollen viability (PV) were investigated in 168 Recombinant Inbred Lines (RIL) and 52 Introgression Lines (IL) derived from the cross between *Solanum lycopersicum* var. “MoneyMaker” and *S. pimpinellifolium* accession TO-937 that were cultivated under increased temperature regimes conditions: T1 (25°C day/21°C night), T2 (30°C day/25°C night) and T3: (35°C day/30°C night). The increase in temperature affected drastically several reproductive traits, for example, FRS in Moneymaker was reduced by 70% and 85% at T2 and T3 when compared to T1 while this reduction ranged from 50% to 90%, in some of the RILs. QTL analysis allowed the identification of genomic regions affecting these traits at different temperatures regimens. A total of 78 QTLs involved in reproductive traits at different temperatures were identified in the RIL population. Most QTLs were temperature range specific except QTLs on chromosomes 1, 2, 6 and 12. Furthermore, ILs with introgressions in chromosomes 1 and 12 maintained good FRN and FRS at T3 in replicated trials. These results represent a catalogue of QTLs and pre-breeding materials that could be used as starting point to decipher the genetic control of the genetic response of reproductive traits at different temperatures and pave the road to develop new cultivars adapted to climate change.

PE0864: Tomato, Potato, Pepper, and related

Development of a TYLCV Diagnostic Marker System in Tomato Utilizing Gene-Specific Markers and Infectivity Assays

Inhwa Yeam¹, Je Min Lee², D Chung² and J Lee², (1)Andong National University, Andong, South Korea, (2)Kyungpook National University, Daegu, South Korea

Tomato Yellow Leaf Curl Virus (TYLCV) is one of the most threatening diseases in tomato production worldwide. TYLCV resistance genes reported to date include *Ty1/3*, *Ty2*, *Ty4*, *ty5* and *Ty6*, all derived from wild-type tomatoes. *Ty1* and *Ty3* originated from *S. chilense* are found to be allelic and encode an RNA-dependent RNA polymerase (RDR) on chromosome 6. *Ty2* derived from *S. habrochaites* encodes a nucleotide-binding domain and leucine-rich repeat containing (NB-LRR) gene on chromosome 11. *Ty4* resistance was reported in *S. chilense* line LA1932 in addition to *Ty3* and shown to increase the level of the resistance. A recessive resistance gene *ty5* originated from *S. peruvianum* is located on chromosome 4 and loss-of-function mutant allele of the *pelota* gene. Another recessive resistance gene *Ty6* derived from LA1938 and LA2779 was recently mapped to chromosome 10. Among all these TYLCV resistance genes known to date, *Ty1* resistance has been most intensively investigated and utilized in the commercial TYLCV resistance breeding programs. However, *Ty1* resistance breaking incidences are continuously reported and achieving durable and stable resistance against TYLCV in the field is acutely needed. In this study, gene-specific markers for *Ty2* and *ty5*, and a closely linked marker for *Ty4* are developed. These newly developed markers in addition to the *Ty1/3* gene specific marker were applied to distinguish TYLCV resistances in various tomato genotypes including our breeding lined and commercial cultivars. Quantitative infectivity assays using both natural infection in the field and artificial inoculation utilizing infectious TYLCV clones in the growth chamber are optimized and performed to precisely address the resistance levels of each resistances and combination of these resistances. Visual symptom development and molecular quantification of relative amount of viral DNA was conducted and analyzed. Exploration of novel resistance sources and analyzing the effect of combining multiple TYLCV resistances is in the process of achieving durable and stable resistance, and we are in the process of pyramiding TYLCV resistances with maximum protection against TYLCV.

PO0865: Tomato, Potato, Pepper, and related

Fine Cis-Engineering By CRISPR/Cas9 in the Promoter of Tomato *KLUH* Generates Varied Fruit Weight for Crop Improvement

Qiang LI, University of Georgia, ATHENS, GA

The current breeding effort and the rate of crop yield increase are likely unable to keep up with the ever-increasing food demands. Genetic diversity is the key source for crop improvement. CRISPR/Cas-mediated *cis*-regulatory region engineering (*cis*-engineering) holds great promise for expanding genetic diversity and accelerating crop improvement. However, current application of CRISPR/Cas-mediated *cis*-engineering with the aim of crop improvement mainly focus on ‘negative regulators’ by creating loss-of-function alleles and has a restricted potential for crop improvement. Moreover, natural genetic variants involved in *cis*-regulation are especially informative and desirable targets of CRISPR/Cas-mediated *cis*-engineering for expanding genetic diversity for breeding. Here, we use CRISPR/Cas9 targeting a natural genetic change, named M9 SNP, in the promoter of tomato *KLUH*, which is a major positive regulator controlling fruit weight. A collection of novel variants encompassing a wide range of fruit weight variation were created in wild-type tomato *LA1589* and semi-cultivated tomato *VIR1011*. Furthermore, 3 conserved motifs including the target site were identified in the promoters of *KLUHs* in other crops, such as potato, pepper, rice, soybean and wheat, providing putative targets for other crops improvements using the approach described in our study, and the beneficial variants identified here can be introduced into elite tomato germplasm by precise genome editing, leading to fast-forward crop improvement.

PO0867: Tomato, Potato, Pepper, and related

Metabolic Interaction between Tomato Scion and Rootstock Under Salt Treatment

Chao Song, Tania Acuna, Michal Meir, Simon Barak and Aaron Fait, Ben-Gurion University of the Negev, Blaustein Insts for Desert Research, Sde Boqer campus, Israel

Salinity is an increasingly expansive problem limiting plant growth and affecting crop production. Grafting is regarded as a promising tool to improve the resistance to soil salinity. Understanding the mechanisms mediating the rootstock affecting the scion is key to broaden the salinity optimum for crops cultivars. In our experiment, we used tomato (*Solanum lycopersicum*) cv M82 grafted onto 254 different tomato rootstocks, including wild and commercial races and cherry tomatoes. Salt treatment was applied at a concentration of 200 mM NaCl and plant tissue was collected in consecutive weeks for metabolite profiling, ROS estimation, transcript analysis. Plant height, branches per plant, fresh weight of shoot and fruit were measured at harvest. Results showed that grafting same scion onto different rootstocks resulted in phenotypic diversification. Compared with M82 self-grafted (control), the MDA content was regarded as the trait with highest CV and standard deviation cross the populations, whereas the ratio of fruit weight to total weight was the trait with the lowest CV. 8 best and 4 worst tomato lines were screened from 254 lines based on fruit fresh weight, harvest index and MDA content in comparison to control. Metabolic data showed that both of proline and serine decreased compared with control cross the populations. Among the screened lines, these best lines accumulated more amino acids, such as alanine, valine and glycine, than worst lines. Further metabolic and transcriptive analysis are being processed to understand the mechanisms that scion was mediated by the rootstock.

PE0868: Tomato, Potato, Pepper, and related

Insights to the Diversity of Cultivated Potato (*Solanum tuberosum*) Using Genotyping-By-Sequencing

Heather Kay Tuttle, University of Minnesota, Minneapolis, MN, Alfonso Del Rio, University of Wisconsin, Madison, WI, John Bamberg, USDA ARS, Sturgeon Bay, WI and Laura M Shannon, University of Minnesota, Saint Paul, MN

Potato is the most widely grown and consumed vegetable crop in the world. Peruvian in origin, it was domesticated from a species complex in the Andes. While wild potatoes are primarily diploid, Peruvian potatoes range from diploid to pentaploid. Cultivated potato in the United States is clonally propagated and consists entirely of highly heterozygous autotetraploids. Reports on the nature and origins of potato diversity, specifically how diversity in landraces and cultivated species relates to diversity in wild species, vary.

Our study aims to provide insights to help resolve these discrepancies, with a collection that represents a cross-section of the US potato genebank in Sturgeon Bay, Wisconsin, which contains 730 individuals consisting of 16 pentaploids, 642 tetraploids, 6 triploids, and 66 diploids. Individuals were genotyped using Genotyping-By-Sequencing (GBS). Genotypes for the diploids were called and hard-filtered using GATK where 258K SNPs, with 18X read depth, were produced with a high amount of missing information. To reduce missingness, genotypes were imputed and 110K SNPs were retained without missing information. Tetraploid genotypes followed the same pipeline and afforded about 12K SNPs with 60X read depth. We examined population structure and diversity measures within and between taxa in the collection.

PO0869: Tomato, Potato, Pepper, and related

Genome Assembly of the Diploid Self Compatible Potato Cultivar Solyntus

Richard Finkers¹, Danny GE Esselink², Fernanda A. de F. Guedes¹, Martijn van Kaauwen¹, Natascha H. M. van Lieshout¹, Pim Lindhout³, Menno Termaat³, Richard G.F. Visser¹, Michiel de Vries³ and Ate van den Burg³, (1)Wageningen University & Research, Plant Breeding, Wageningen, Netherlands, (2)Wageningen UR, Wageningen, Netherlands, (3)Solynta, Wageningen, Netherlands

The potato is one of the most important food crops worldwide. Improvements to its traits can therefore have a major impact. Reading the genome structure of the potato is extremely tricky, however, as a regular potato consists of four genomes, which makes it difficult to determine the position of the genes. The recent research applied a diploid real potato plant with only one genome, a so-called homozygote, which makes it easier to read and compare the DNA base sequence. This line, Solyntus, was produced as part of Solynta's hybrid potato breeding programme.

The previously available genome sequence, which I also helped establish, consisted of approximately 125,000 small segments. The genome we are presenting now comprises 185 large segments. This is a significant improvement which was achieved via a combination of unique plant material and new sequencing and analysis techniques. While the previous sequence involved a wild variety of the potato, we have now used an actual potato plant. To stimulate research on diploid potato, we are now releasing a good performing inbred line as a universal research line. Solyntus, shows vigorous growth combined with high homozygosity. In greenhouse experiments, good tuber yields and numbers were obtained. It produces round tubers with creamy. Next to fertility and other characteristics of Solyntus, we will show the level of homozygosity, assembly statistics as well as a first comparison to the DM reference genome.

Solyntus, and the genome sequence, is available for scientific research and is made available for the scientific community via the website: www.plantbreeding.wur.nl/Solyntus/.

PE0870: Tomato, Potato, Pepper, and related

Comparing the Core and Accessory Genome of Diploid Potato

Maria Kyriakidou¹, Helen H. Tai², Noelle Anglin³, David Ellis³ and Martina Stromvik⁴, (1)Department of Plant Science - McGill University, Montreal, QC, Canada, (2)Agriculture and Agri-Food Canada, Fredericton, NB, Canada, (3)International Potato Center, Lima, Peru, (4)Department of Plant Science - McGill University, Ste. Anne de Bellevue, QC, Canada

Potato (*Solanum tuberosum* L.) is an important staple crop, cultivated at various altitudes and climates globally. Increasing resources beyond a single reference genome is needed to capture the great diversity that enables this crop to be so successful and to grow in various environments. In this work, we have sequenced the genomes of six native Peruvian diploid potato accessions to construct a pan-genome. Next Generation (NGS) and Third Generation Sequencing (long and linked reads) data have been generated for the *de novo* assembly of these genomes. The newly diploid assembled genomes (based on Hawkes taxonomy - *S. stenotomum* subsp. *stenotomum*, *S. phureja*, *S. xajanhui*, *S. bukasovii*, plus two accessions from *S. stenotomum* subsp. *goniocalyx*), along with the genomes of two wild, publicly available diploid potato genomes; *S. commersonii* and *S. chacoense* M6 clone were compared to the DM1-3 reference genome to construct a diploid potato pan-genome. We predicted 723 coding genes not present in the DM1-3 reference genome. Approximately, 71% of the genes consist of the core genome (present in all the genomes), while 29% consists of the accessory genome. Within the accessory genome, 1.39% of the genes are

present only in the DM1-3, but not in the rest of the genomes. Moreover, another 1.37% of the accessory genome consist of genome-specific genes. Gene ontology functional analyses indicate that the newly identified protein coding genes including those associated with disease resistance and self-incompatibility.

PO0871: Tomato, Potato, Pepper, and related

Exploring Genome Diversity in Potatoes of Various Ploidy Levels

Maria Kyriakidou¹, Saireddy Achakkagari², Jose Hector Galvez², Xinyi Zhu³, Kevin Tang², Helen H. Tai⁴, Noelle Anglin⁵, David Ellis⁵ and Martina Stromvik², (1)Department of Plant Science - McGill University, Montreal, QC, Canada, (2)Department of Plant Science - McGill University, Ste. Anne de Bellevue, QC, Canada, (3)McGill University, Montreal, QC, Canada, (4)Agriculture and Agri-Food Canada, Fredericton, NB, Canada, (5)International Potato Center, Lima, Peru

Whole genome sequencing has allowed the expansion of genomic resources for the exploration of important agronomic traits in potato. An important source of plant genome diversity is copy number variation (CNV). Studies done by other researchers have demonstrated high frequency of CNV events in potato mostly in Phureja and Stenotomum diploids groups. This study used whole genomic sequences of six diploid (based on Hawkes taxonomy - *S. stenotomum subsp. stenotomum*, *S. phureja*, *S. xajanhui*, *S. bukasovii*, plus two accessions from *S. stenotomum subsp. goniocalyx*), two triploid (*S. juzepczukii*, *S. chaucha*), three tetraploid (*S. tuberosum subsp. andigena*; two accessions, *S. tuberosum subsp. tuberosum*) and a pentaploid (*S. curtilobum*) potato landraces, in addition to two wild, diploid publicly available genomes (*S. commersonii* and *S. chacoense*) to explore structural variation compared to two potato reference genomes; DM1-3 and *S. chacoense* M6. In agreement with previous work, the significant gene CNV clusters include genes involved in metabolic processes, disease resistance and stress tolerance. The CNV-status of each gene in these genomes was used for the classification of these potato genomes. Finally, a single nucleotide analysis (SNP) uncovered the regions of heterozygosity in these genomes.

PE0872: Tomato, Potato, Pepper, and related

Genetic Structure and Molecular Diversity Among Clones Representing the Texas A&M Potato Breeding Program

Jeewan Pandey¹, Douglas C. Scheuring¹, Jeffrey W. Koym², Angel Chappell¹, Julian Creighton Miller¹, Joseph Coombs³, David Douches³ and Maria I. Vales¹, (1)Texas A&M University, College Station, TX, (2)Texas A&M AgriLife Research and Extension Center, Lubbock, TX, (3)Michigan State University, East Lansing, MI

The Texas A&M Potato Breeding and Variety Development Program has practiced the introduction of early generation experimental selections into *in-vitro* culture due to the absence of seed production and certification programs in Texas. More than 200 experimental clones and varieties have been preserved in tissue culture. These clones represent our main breeding gene pool. The extent of genetic variability present in the breeding pool and the presence of major genes of interest are unknown. Evaluating the genetic diversity of this gene pool is essential to understanding its genetic structure. Resources are now available to better characterize germplasm diversity in breeding programs and to accelerate selection by using genomics-enabled approaches. So far, 218 clones have been genotyped using the Infinium 22K V3 Potato Array. Tetraploid potatoes, representing different market classes were included in this study. Hierarchical clustering, based on 10,698 polymorphic SNP loci, was performed using the R package 'ape'. In general, accessions clustered together based on market class, known provenance, and ploidy level. As expected, the level of heterozygosity in tetraploid clones was much high (average 61%). Population structure and discriminant analysis of principal components (DAPC) grouped the population into three well-defined clusters. The clusters were associated with different potato market classes. Ultimate goal of this work is to genetically characterize the entire collection of breeding clones, conduct marker-trait association studies and guide future crossing schemes.

PO0873: Tomato, Potato, Pepper, and related

Genetic Mapping of the *c1* Locus By GBS-Based BSA-Seq Revealed *Pseudo-Response Regulator 2* As a Candidate Gene Controlling Pepper Fruit Color

Soo Bin Lee, Jeong Eun Kim, Gyu-Myung Lee and Je Min Lee, Kyungpook National University, Daegu, South Korea

The loci *c1*, *c2*, and *y* have been widely reported as genetic determinants of various ripe fruit colors in pepper. However, *c1*, which may impact reduced pigmentation in red, orange, and yellow fruits, is not well understood. Two cultivars showing peach or orange fruit in *Capsicum chinense* ‘Habanero’ were found to have *c2* mutation and were hypothesized to segregate *c1* locus in the F₂ population. Habanero peach (HP) showed a reduced level of chlorophylls, carotenoids and total soluble solids in immature and ripe fruit. A microscopic examination of the fruit pericarps revealed smaller chloroplasts and less stacked thylakoid grana in HP. The expression of many genes related to chlorophyll and carotenoid biosynthetic pathways were reduced in HP. To identify the genomic region of the *c1* locus, bulked segregant analysis combined with genotyping-by-sequencing was employed on a F₂ population derived from a cross between Habanero orange and HP. One SNP at chromosome 1 was strongly associated with the peach fruit color. Pepper *Pseudo-Response Regulator 2* (*PRR2*) was located close to the SNP and cosegregated with the peach fruit color. A 41 bp deletion at the third exon-intron junction region of *CcPRR2* in HP resulted in a premature termination codon. A nonsense mutation of *CaPRR2* was found in *C. annuum* ‘IT158782’ which had white ripe fruit coupled with null mutations of *capsanthin-capsorubin synthase* (*y*) and *phytoene synthase 1* (*c2*). These results will be useful for the genetic improvement of fruit color and nutritional quality in pepper.

PE0874: Tomato, Potato, Pepper, and related

Light-Dependent Regulation of Fruit Pigments and Volatile Flavors in Bell Pepper and Tomato Under an Artificial Canopy

Eun Sol Kang, Hee Ju Yoo and Je Min Lee, Kyungpook National University, Daegu, South Korea

Light is a major environmental factor in the regulation of secondary metabolisms such as pigments and flavors. To investigate the light-dependent regulation underlying fruit pigmentation and volatile flavors, major fruit pigments, their biosynthetic gene expression, and volatile flavors were analyzed in tomato and bell pepper fruit under a fruit canopy. Immature canopy-treated fruit was ivory in color and no chlorophyll was detected in either plant. Carotenoid and flavonoid levels were also reduced in ripe tomato and bell pepper fruit under the canopy. Light positively impacts fruit pigmentation in tomato and bell pepper by regulating gene expression in carotenoid and flavonoid biosynthesis, especially *PHYTOENE SYNTHASE* and *CHALCON SYNTHASE*, respectively. Interestingly, the fruit canopy promoted ripening. Nineteen volatile flavors were detected in both ripe tomato and pepper fruits, and seven of these exhibited light-dependent regulation. This study will be helpful for understanding the molecular mechanisms and improving fruit quality in the biosynthesis of pigments and flavors.

PO0875: Tomato, Potato, Pepper, and related

Characterization of Secreted Proteins of *Colletotrichum scovillei* during Pepper Anthracnose Disease Development

Kyoung Su Kim, Kangwon National University, Chuncheon, South Korea

Colletotrichum scovillei is an ascomycete plant pathogen responsible for pepper anthracnose, the major problem in many countries over the world. *C. scovillei* is a member in *C. acutatum* species complex. A group of small secreted proteins was identified via combined bioinformatics approach from the genome of *C. scovillei*. *C. scovillei* develops appressorium on pepper fruit and penetrates cuticle layer of pepper fruit. A highly branched dendroid structure is formed in the cuticle layer of pepper fruit. We observed thin and branched hypha inside dendroid structure that grew along with dendroid structure toward the wall of the infected cell. To visualize small secreted proteins of *C. scovillei* during disease development on pepper fruit, a total of 39 proteins were tagged with GFP. We also generated many deletion mutants with homology-dependent targeted gene replacement. In this presentation, patterns of cellular localization of small secreted proteins will be presented, which indicate their potential roles in establishing disease development. Also, phenotypic characterization of deletion mutants revealed that corresponding genes are involved in not only disease development, but also the fungal development. Collectively, our results provide fundamental basis in understanding pathogenic development of *C. scovillei* during anthracnose disease, which can be useful in anthracnose management of many crops.

PE0876: Tomato, Potato, Pepper, and related

The Ankyrin Repeat Gene Family in *Capsicum* spp: Genome-Wide Survey, Characterization and Gene Expression Profile

Carlos Ortiz¹, Yadira Peña-García¹, Purushothaman Natarajan¹, Menuka Bhandari¹, Lakshmi V. Abburi¹, Lav Yadav¹, John Stommel², Padma Nimmakayala¹ and Umesh K. Reddy¹, (1)Department of Biology, West Virginia State University, Institute, WV, (2)Genetic Improvement of Fruits and Vegetables Laboratory, Beltsville, MD

The ankyrin (ANK) repeat protein family is largely distributed across plants and has been found to participate in multiple processes such as plant growth and development, hormone response, response to biotic and abiotic stresses. It is considered as one of the major markers of capsaicin content in pepper fruits. In this study, we performed a genome-wide identification and expression analysis of genes encoding ANK proteins in three *Capsicum* species: *Capsicum baccatum*, *Capsicum annuum* and *Capsicum chinense*. We identified a total of 87, 85 and 96 ANK genes in *C. baccatum*, *C. annuum* and *C. chinense* genomes, respectively. Next, we performed a comprehensive bioinformatics analysis of the *Capsicum* ANK gene family including gene chromosomal localization, Cis-elements, conserved motif identification, intron/exon structural patterns and gene ontology classification as well as profile expression. Phylogenetic and domain organization analysis grouped the *Capsicum* ANK gene family into eight subfamilies distributed across all 12 pepper chromosomes at different densities. Analysis of the expression of ANK genes in leaf and pepper fruits suggested that the ANKs have specific expression patterns at various developmental stages in placenta tissue. Our results provide valuable information for further studies of the evolution, classification and putative functions of ANK genes in pepper.

PO0877: Tomato, Potato, Pepper, and related

Integrated Transcriptomic and Metabolomic Analysis to Characterize the Cutin Biosynthesis between Low and High Cutin Genotypes of *Capsicum chinense* Jacq.

Purushothaman Natarajan¹, Umesh K. Reddy¹, Carlos Ortiz¹, Padma Nimmakayala¹, John Stommel², Thangasamy Saminathan¹ and Tolulope Akinmoju¹, (1)Department of Biology, West Virginia State University, Institute, WV, (2)Genetic Improvement of Fruits and Vegetables Laboratory, U.S. Department of Agriculture, Agricultural Research Service, Beltsville Agricultural Research Center, Beltsville, MD 20705-2325, USA, Beltsville, MD

Terrestrial plants are constantly faced with biotic and abiotic stresses such as pathogen/pest infections, extreme temperatures, drought, and UV radiation. As a result of this, plants have developed several mechanisms to tolerate these stresses. One such trait, the cuticle plays crucial roles in plant survival and physiology. In this study, we have integrated the metabolome and transcriptome profiling pertaining to cutin in two habanero genotypes; PI 222448 and PI 257145. The fruits were selected based on the waxy or glossy phenotype on their surfaces. Metabolomics analysis using GC-MS, showed a significant variation in the cutin composition across two genotypes with about 6-fold higher cutin content in PI 257145. The GC-MS analysis also revealed that the 10,16-dihydroxy hexadecanoic acid monomer was present at the highest percentage (82.6%) in PI 257145. Transcriptomic analysis of high (PI 257145) and low cutin (PI 222448) genotypes using RNA-Seq resulted in identification of 2,703 statistically significant differentially expressed genes (DEGs) which includes 1,693 genes up-regulated in high cutin genotype PI 257145, while 1010 genes were down-regulated. Gene annotation, GO analysis, pathway and network analyses of the DEGs revealed enriched genes and pathways that are differentially expressed between the two genotypes. Our study showed that the many key genes involved in the cutin biosynthesis were differentially expressed in several folds between the low and high cutin genotypes. Cutin synthase, glycerol-3 phosphate acyltransferase 6, long-chain acyltransferase 2, long-chain acyltransferase 1 and CYP86A4 were found to be expressed at higher level in PI 257145, which were also consistent with the metabolome data. Details of the study will be presented.

PE0878: Tomato, Potato, Pepper, and related

Acid Mitigating Mechanisms in *Capsicum* Spp.

Tosin Akinsipe, Lakshmi V. Abburi, Suhas Shinde, Thangasamy Saminathan, Padma Nimmakayala and Umesh K. Reddy, Department of Biology, West Virginia State University, Institute, WV

The distribution of major nutrients in soil is determined by the pH and an imbalance can impede absorption of nutrients by plants. In this study, phenotypic variation of root architecture was used to identify genes associated with the tolerance or susceptibility to low pH among three cultivated species of peppers. Population containing 116, 113 and 104 lines of *Capsicum baccatum*, *C. chinense* and *C. annuum* respectively, were grown in germination pouches with reduced pH (3.0) of nutrient solution. Automated phenotype software was used to determine root length, surface area, average diameter, root volume, tips, forks and crossings. GWAS was conducted for all the root traits in various panels separately using 20,312 SNPs, of which 3875 SNPs were polymorphic in *C. baccatum*, 7384 for *C. chinense* and 4065 for *C. annuum*. Our GWAS resulted in 850 genes across the 12 chromosomes. Our simultaneous GWAS analysis of Arabidopsis identified several common candidates that were not previously known to be important for the acid tolerance. We identified 175 common genes between *Capsicum* species and Arabidopsis associated for acidity and Aluminum tolerance. Nitrate transporter1/Peptide transporter, NTR1/PTR Family 5.6-like, NAC transcriptional factor 25, CAX18 Cation/H⁺ antiporter 18-like and Pyrophosphate-energized vacuolar membrane proton pump were some of the important genes currently under investigation.

PO0879: Tomato, Potato, Pepper, and related

Characterization of Sweetpotato Inheritance Using Ultradense Multilocus Genetic Map

Marcelo Mollinari¹, Bode Adebawale Olukolu², Guilherme Da Silva Pereira¹, Awais Khan³, Dorcus C. Gemenet⁴, G. Craig Yencho⁵ and Zhao-Bang Zeng¹, (1)North Carolina State University, Raleigh, NC, (2)University of Tennessee, Knoxville, TN, (3)Plant Pathology and Plant-Microbe Biology, Cornell University, Geneva, NY, (4)International Potato Center (CIP), Lima, Peru, (5)North Carolina State University (NCSU), Raleigh, NC

The cultivated sweetpotato (*Ipomoea batatas* (L.) Lam., 2n = 6x = 90) is an important staple food crop with an annual production of 112.84 million tons. Despite its undeniable social and economic importance, genetic studies in sweetpotato significantly lag behind major diploid crops due to its complex polyploid genome. To fully characterize the inheritance pattern in sweetpotato, we built an ultra-dense multilocus integrated genetic map of a full-sib population derived from a cross between the cultivars ‘Beauregard’ and ‘Tanzania’ (BT population) using our newly implemented software, MAPpoly. The resulting genetic linkage map consisted of 30,684 SNPs distributed in 15 homology groups with a total length of 2708.3 cM (11.3 SNPs/cM). We observed 96.5% collinearity between *I. batatas* and its diploid relative *I. trifida*. Using the genotypic probabilities computed across all linkage groups, we inferred the complete hexaploid haplotypes for all individuals in the offspring. We also observed that 73.3% of the meiotic configurations in the parents were resolved in bivalents, 15.7% presented multivalent signatures, and 11.0% were inconclusive. Moreover, the BT population presented vastly hexasomic inheritance mechanisms in all linkage groups, except for linkage group 2, which presented low levels of preferential pairing in parent Tanzania. We propose that the hexasomic-bivalent inheritance promotes stability to the allelic transmission in sweetpotato.

PE0880: Tomato, Potato, Pepper, and related

Contrasting N-Use Efficient Eggplants and Differential Expressed Genes in the N-Metabolism Pathway

Antonio Mauceri¹, Antonio Lupini¹, Laura Toppino², Maria Rosa Abenavoli¹, Giuseppe Leonardo Rotino² and **Francesco Sunseri**¹, (1)University Mediterranea of Reggio Calabria, Reggio Calabria, Italy, (2)CREA-GB, Montanaso Lombardo, Italy

Nitrogen availability is one of the most limiting factors affecting crop yield. The identification of high Nitrogen Use Efficiency (NUE) genotypes may represent a valuable strategy to maintain high yield reducing N-supply. The genetic variation among eggplant genotypes in response to nitrate supply was exploited allowing us to identify two pairs of NUE-contrasting genotypes. RNAseq analysis was then assessed to identify differentially expressed genes (DEG) related to NUE pathway, after exposure to short- and long-term N-stress.

DEGs identification between genotypes in responses to low N supply were taking into account and comparisons were independently analyzed in both root and shoot. In particular, four up regulated genes were identified in high NUE genotypes by co-expression networks (GCN) analysis. Transcriptomics analysis highlighted the central role of some transcription factor (TF), which were up regulated in the N-use efficient genotypes. A TF belonging to WRKY

family, involved in plant stress responses, showed a significant up-regulation. Interestingly, this WRKY33 triggered a higher expression of 21 genes including other TFs, many of which closely related to N-assimilation and remobilization.

These results fit well with the evidences of the key role of N-utilization component (NUE) to confer high NUE in eggplant suggesting the higher N-remobilization to the fruit, driven by GS enzyme, as a valuable strategy to enhance NUE.

PO0881: Tomato, Potato, Pepper, and related

Downregulation of Imidacloprid Resistant Genes Leads to Reduced Fecundity and Survival of Colorado Potato Beetle, *Leptinotarsa decemlineata* (Chrysomelidae: Coleoptera)

Muhammad Nadir Naqqash, Ayhan Gökçe and Allah Bakhsh, Nigde Omer Halisdemir University, Nigde, Turkey

Colorado potato beetle, *Leptinotarsa decemlineata* Say (coleoptera: chrysomelidae), is the important pest of potato all over the world. This insect pest is resistant to more than 50 active compounds belonging to various chemical groups. Potential of RNA interference (RNAi) was explored to knock down transcript levels of imidacloprid resistant genes in Colorado potato beetle (CPB) under laboratory conditions. Three important genes belonging to cuticular protein (CP), cytochrome P450 monooxygenases (P450) and glutathione synthetase (GSS) families encoding imidacloprid resistance were targeted. Feeding bioassays were conducted on various stages of imidacloprid resistant CPB lab population by applying HT115 expressing dsRNA on potato leaflets. Survival rate of insects exposed to CP-dsRNA decreased to 4.23%, 15.32% and 47.35% in 2nd, 3rd and 4th instar larvae respectively. Larval weight and pre-adult duration were also affected due to dsRNAs feeding. Synergism of RNAi with imidacloprid conducted on the 2nd instar larvae, exhibited 100% mortality of larvae when subjected to reduced doses of GSS and CP dsRNAs along with imidacloprid. Utilization of three different dsRNAs against imidacloprid resistant CPB population reveal that dsRNAs targeting CP, P450 and GSS enzymes could be useful tool in management of imidacloprid resistant CPB populations.

PE0882: Tomato, Potato, Pepper, and related

WAT1 Determines the High Levels of Auxin Signalling in Cambium Stem Cells during Wood Formation

Hojin Ryu, Chungbuk National University, Cheongju, South Korea

The crosstalk of plant hormones at multiple levels serves as a robust regulatory node responsible for the optimal output of cellular processes through the integration of endogenous and exogenous signalling cues . Xylem differentiation and wood formation during the secondary growth of plants entirely depend on unique combinations of plant hormone signals that involve auxin, cytokinins, abscisic-acid, gibberellins, and brassinosteroids in a spatiotemporal manner. Specifically, the tight regulation of local auxin homeostasis and formation of a signalling maximum in xylem precursor cells specify the organizing activity that balances the bidirectional cell proliferation and differentiation of conducting cells from bifacial cambial stem cells. However, the molecular basis underlying the local control of auxin accumulation and subsequent signalling maximum remains unknown. Here, we reveal that SIWAT1 (Walls Are Thin1) facilitates xylem differentiation and wood formation by specifying the stem cell organizing activity of the vascular cambium in tomato (*Solanum lycopersicum*). The combination of genetic manipulation using molecular toolkits and histological analysis of gain- and loss-of-function mutants revealed that SIWAT1 is a key component of auxin distribution under the control of BR signalling for xylem differentiation in the tomato stem and root. The direct activation of local auxin responses by upregulation of *SIWAT1* via BR-activated SIBZR1/2 in cambial cells adjacent to xylem cells determines the stem cell organizers of the vascular cambium. Our data suggest that the SIBZR1/2-SIWAT1 signalling nexus dictates the high levels of auxin signalling in vascular cambium organizers for wood formation.

PO0883: Wheat, Barley, Oat, and related

The International Wheat Genome Sequencing Consortium (IWGSC)

The International Wheat Genome Sequencing Consortium¹, Rudi Appels², Ute Baumann³, Hikmet Budak⁴, Isabelle Caugant¹, Jan Dvorak⁵, **Kellye Eversole**¹, Ming-Cheng Luo⁶, Etienne Paux⁷ and Sébastien Prud'homme⁸, (1)IWGSC, Lee's Summit, MO, (2)University of Melbourne, Melbourne, VIC, Australia, (3)The University of Adelaide, Urrbrae, South Australia, Australia, (4)Montana BioAgriculture Inc., Missoula, MT, (5)Department of Plant Sciences, University of California, Davis, CA, (6)University of California, Davis, CA, (7)INRA GDEC, Clermont-Ferrand, France, (8)LIMAGRAIN, CHAPPES, France

Bread wheat, the staple food for 35% of the world's population, is the last major crop species to benefit from a reference genome sequence. The IWGSC, with 2,400 members in 68 countries, is an international, collaborative consortium, established in 2005 by public and private wheat growers, breeders and scientists, with the aim of delivering genomic tools and resources for wheat improvement.

In 2018, the IWGSC completed Phase I when it published the first high quality reference sequence of the bread wheat variety *Chinese Spring* (IWGSC RefSeq v1.0). The IWGSC RefSeq v1.0 represents 94% of the hexaploid wheat genome in 21 chromosome pseudomolecules; and identifies 107,891 high confidence genes, along with 4.7 million molecular markers.

The IWGSC has now entered a new phase of its activities. In particular, an improved version of the reference sequence, IWGSC RefSeq v2.0, which closed a number of gaps and corrected position and orientations of scaffolds, was released to the community in July 2019 under the Toronto pre-publication access agreement. The development of annotation v2.0 integrating functional and manual annotation, as well as alignment with other genomic resources, is also in progress. Finally, a wheat diversity project has begun, aiming at completing high quality sequences of landraces representing the worldwide genetic diversity. These activities lay the foundation for genomics-based wheat improvement in response to challenges imposed by population expansion and climate change.

An overview of the IWGSC recent achievements will be presented, along with an outline of its current activities.

PE0884: Wheat, Barley, Oat, and related

Building a Pan-Transcriptome of Wheat

Benjamin H White¹, Susan Duncan¹, Hannah Rees¹, Rachel Rusholme Pilcher¹, Anthony Hall¹ and The 10 Plus Wheat Genomes Project, (1)Earlham Institute, Norwich, United Kingdom

With more than 215 million hectares planted annually, wheat is the most widely cultivated cereal in the world. In order to keep up with the demands of a growing global population, wheat production needs to increase by around 60% within the next 40 years. The challenges for wheat breeders and growers are, therefore, tremendous. With a complete reference bread wheat genome, the international community is sequencing and assembling more wheat varieties in order to quantify diversity and identify genomic regions under selection (human and natural selection). The 10 plus wheat genome project has generated do novo assembly of 16 elite wheat genomes. With these new data sets we can start to ask functional genomic question of how differences in the genomes lead to differential expression patterns between cultivars and how breeding and domestication have driven changes in their transcriptome. Analysis of this “pan-transcriptome”, as part of this global partnership, will provide a unique understanding of how gene-expression changes between these lines and will characterise core and dispensable genes; those expressed all lines or in only one. It will give a unique understanding of how transcription networks change between cultivars and the impact of having three genome giving both flexibility and robustness to transcription networks. As such, the resources developed from this will be directly applicable to the community and breeders towards the identification of genes or network changes that can be used to improve key traits, such as increased yield and disease resistance. In addition it will improving our fundamental understanding of gene models and regulatory motifs in wheat.

PO0885: Wheat, Barley, Oat, and related

Global Durum Wheat Panel: A Genetic Tool for Durum Wheat Diversity Studies

Luigi Cattivelli¹, Elisabetta Mazzucotelli¹, Marco Maccaferri², Anna-Maria Mastrangelo³, Giuseppe Sciarra⁴, Steven Xu⁵, Justin Faris⁶, Matthew Hayden⁷, Penny J. Tricker⁸, Hakan Ozkan⁹, Viviana Echenique¹⁰, Brian J. Steffenson¹¹, Ron E. Knox¹², Sripada Udupa¹³, Friedrich Longin¹⁴, Daniela Marone³, Giuseppe Petruzzino³, Aziz Niane¹⁵, Pablo Roncallo¹⁶, Ahmed Amri¹³, Hans-Joachim Braun¹⁷, Karim Ammar¹⁷, Michael Baum¹³, Roberto Tuberosa¹⁸ and Filippo Maria Bassi¹³, (1)CREA - Research Centre for Genomics and Bioinformatics, Fiorenzuola d' Arda, Italy, (2)Department of Agricultural and Food Sciences (DISTAL), University of Bologna, Bologna, Bologna, Italy, (3)CREA Research centre for cereal and industrial crop, Foggia, Italy, (4)Department of Agricultural and Food Sciences - University of Bologna, Bologna, Italy, (5)USDA-ARS, Edward T. Schafer Agricultural Research Center, Fargo, ND, (6)USDA-ARS, Fargo, ND, (7)Agriculture Victoria, AgriBio, Centre for AgriBioscience, Melbourne, Vic, Australia, (8)School of Agriculture, Food and Wine, University of Adelaide, Adelaide, SA, Australia, (9)University of Cukurova, Adana, Turkey, (10)Centro de Recursos Naturales Renovables de la Zona Semiárida (CERZOS) Universidad Nacional del Sur-CONICET, Bahía Blanca, Argentina, (11)Department of Plant Pathology, University of Minnesota, Saint Paul, MN, (12)Swift Current Research and Development Center, Swift Current, SK, Canada, (13)ICARDA, Rabat, Morocco, (14)University of Hohenheim, Stuttgart, Germany, (15)ICARDA, Beirut, Lebanon, (16)CERZOS CCT - CONICET, Bahía Blanca, Argentina, (17)CIMMYT, Mexico D.F., Mexico, (18)Department of Agricultural and Food Sciences, University of Bologna, Bologna, Italy, Bologna, Italy

A Global Durum Wheat Panel (GDP) of 1,011 genotypes was established through a cooperative effort of durum wheat scientists within the framework of the Expert Working Group on Durum Wheat Genomics and Breeding, part of the Wheat Initiative. Following extensive genotyping with the wheat iSelect 90K SNP assay and a process of selection, nearly 20,000 unique, non-redundant, single Mendelian SNP markers that were both genetically and physically mapped were used to characterise the GDP for linkage disequilibrium (LD), genetic diversity, population structure and genetic relationships. According to the analysis with Admixture, grouping statistics stabilized at $K > 11$. A refined analysis was carried out on 473 durum wheat cultivars representing 12 breeding programs from all around the world. Following an analysis of molecular variance, the highest proportion of variance was observed within groups ($> 88\%$) partitioned based on breeding group and year of release. Moderate levels of genetic diversity were observed for the breeding programs considered, ranging from H_e 0.30 for ICARDA and CIMMYT groups up to H_e 0.41 for the Ethiopian modern germplasm. A moderate trend of decrease in genetic diversity (H_e from 0.38 to 0.34) was observed with respect to the year of release, across groups of cultivars released in decades from 1970 to 2018. Finally, four genome-wide selection indexes (F_{st} , DRI, XP-EHH and XP-CLR) estimated chromosome regions putatively subjected to selection in comparisons of both landrace *vs.* cultivar subpopulations and between cultivar subpopulations.

PE0886: Wheat, Barley, Oat, and related

Genome Editing of Wheat Grain-Regulatory Genes for Novel Variation to Increase Wheat Genetic Yield Potential

Ajay Gupta, South Dakota State University, Brookings, SD

To feed the growing population, wheat yield needs to be increased by 50% in 20 years, which demands an increase of the annual genetic yield gain from $\sim 1\%$ to 1.7%. This quantum gain in genetic yield will require the development of breakthrough approaches and novel genetic resources for wheat improvement. This NIFA-IWYP project aims to develop an improved CRISPR/Cas9 system, create edit mutations for grain-size and grain-number candidate genes, and characterize the phenotype effect of the mutations. We have developed an *Agrobacterium*-delivered CRISPR/Cas9 system with two types of guide RNA cassettes, one for targeting single genes with two guide RNAs and the second one for targeting multiple genes with up to eight guide RNAs. An advantage of these *Agrobacterium*-delivered systems is a requirement of a small number of transformation events to recover desired mutations in T1 or T2 generations. Using these CRISPR/Cas9 systems, we have generated 17 CRISPR constructs targeting 16 genes for wheat transformation, obtained >80 mutations derived from six constructs for five grain-regulatory genes, 63% of which are due to deletions larger than 20 bp. We have characterized five mutations from the *TaCKX2-1*, *TaGLW7*, and *TaGW2* for grain number and grain size. Under the greenhouse conditions, a 1,160-bp deletion in *TaCKX2-D1*

and 5-bp deletion in *TaGLW7-D* could increase grain number up to 140% and 121%, respectively. A 10-bp deletion in *TaGLW7-A*, a 17-bp deletion in *TaGW2-A*, and a 1-bp deletion in *TaGW2-D* increased grain-size to 108%, 107%, and 113%, respectively. We are combining the mutations from homoeologous loci and selecting transgene-free mutants, which can be used as novel germplasm for breeding high grain yield.

PO0887: Wheat, Barley, Oat, and related

Exploring the Genotypic Profile of Alleles Associated with Yield in Selected Wheat Genotypes

Timmy A Baloyi¹, PE Kyeswa¹, Tebogo Oliphant², Tsepiso Hlongoane³ and Toi John Tsilo²,
(1)Agricultural Research Council, Bethlehem, South Africa, (2)Agricultural Research Council - Small Grain Institute, Bethlehem, South Africa, (3)Agricultural Research Council - Small Grains, Bethlehem, South Africa

Breeding wheat (*Triticum aestivum* L.) with high yielding potential has been the ultimate goal of many research and breeding programmes around the world. This is because producers desire high yielding genotypes for profitability. Grain yield is a complex trait which is influenced by a number of (a) biotic factors and various traits in wheat. These include, but not limited to plant height, flowering time, vernalisation, thousand kernel mass, disease resistance and abiotic stress tolerance.

There are number of genes that have been widely reported to have a major influence on grain yield. Amongst others include genes for plant height (Rht-B1, Rht-D1, Rht8), photoperiod response (Ppd-D1) and vernalisation (Vrn-1) genes. Rht-B1b and Rht-D1b alleles that belong to the Rht genes result in the reduction in plant height by conferring the plants insensitive to gibberellic acid (GA) as opposed to their wild-type alleles (Rht-B1a, Rht-D1a) found in tall plants. A reduction in plant height improves lodging resistance, the partitioning of assimilates and ultimately increases grain yield (Ellis et al., 2002, Wilhelm et al., 2013). The Ppd-D1a allele has also been associated with improved grain yield through the reduction in height and the number of days to heading as opposed to the photoperiod sensitive allele Ppd-D1b (Beales et al., 2007; Wilhelm et al., 2013; Worland et al., 2001). The presence of the Vrn-A1 allele provides complete insensitivity to vernalisation, whereas Vrn-B1 and Vrn-D1 alleles result in the reduction in vernalisation requirement (Kolev et al., 2010). Other puriondoline allele variants such as the Pinb-B2v3 allele have also been associated with an improvement in yield due to the presence of desirable grain yield components when detected in wheat genotypes that contain this allele than in those that contain the Pinb-B2v2 allele (Chen et al., 2010; Wilkinson et al., 2008).

The development of high yielding wheat genotypes remains the major objective of South African breeding programmes. Thus, the genotypic profile of traits associated with yield was explored. Whereby yield components of diverse wheat genotypes that are used for pre-breeding have been screened with various molecular markers to evaluate their influence on selected yield components.

PE0888: Wheat, Barley, Oat, and related

Genome Wide Analysis and Prediction of Yield Traits in Soft Red Winter Wheat

Dylan L. Larkin, University of Arkansas, Fayetteville, AR

Wheat (*Triticum aestivum* L.) is a widely produced grain crop, globally, significantly contributing to food security. As populations increase, so will food demands. In order to meet such demands, breeders must increase wheat yield potential. Yield is impacted by multiple quantitative traits, relying on several quantitative trait loci (QTL). A genome-wide association study (GWAS) was conducted, using 354 soft red winter wheat (SRWW) genotypes adapted to the southern United States in an association mapping panel (AMP) to identify novel QTL associated with five quantitative traits, including grain yield (GY), test weight (TW), heading date (HD), maturity date (MD), and plant height (PH). The AMP was grown over eight location-years between 2013 and 2017 in randomized complete block and augmented designs. Each location-year was evaluated for the five aforementioned traits. Adjusted means for each trait were obtained from a random field linear mixed model for each location-year and combined to obtain best linear unbiased predictions from SAS 9.4 software. Marker-trait associations (MTA) were determined using FarmCPU in R v3.6.1 software. Highly significant MTA (LOD>6.15) were observed on 13 of 21 chromosomes for all five traits. A genomic prediction cross-validation analysis was also performed for all five traits using significant

MTA from the GWAS as fixed effects to determine if they could improve prediction accuracy. Results from this study will facilitate the development of SRWW cultivars with greater yield potential.

PO0889: Wheat, Barley, Oat, and related

Genetic Analysis of Allelic Variation in Heading Date of Hexaploid Wheat Varieties

Fatma Gul Maras Vanlioglu, Republic of Turkey Ministry of Agriculture and Forestry, Ankara, NC, Turkey and **Zemran Mustafa**, Republic of Turkey Ministry of Agriculture and Forestry, Ankara, Turkey

The sessile lifestyles of plants force them to adapt to the changes in the environment. The plants especially wheat (*Triticum aestivum* L.) which has the largest cultivation area in the world, is one of the major staple food. It can be grown in a wide range of climate conditions from cold high latitudes to hot equatorial regions and they have been adapted to grow in many different conditions with breeding techniques. One of the most severe effects on plants is winter injury at the cold climates. Thus, the studies about understanding the behavior of cold-tolerant wheat varieties are crucial. It is known that vernalization and photoperiod genes influence the heading date (HD) by controlling the transition from vegetative to the reproductive stage which protect floral meristem from the effects of winter. The allelic variation in vernalization may change the requirement of cold treatment for spike primordia differentiation. In addition to vernalization, the variation in photoperiod loci may affect the sensitivity of plants to the day-length. Both vernalization requirements and photoperiod sensitivity can provide low-temperature tolerance by keeping plants in the vegetative state during cold winter days. In this study, 40 different hexaploid wheat varieties cultivated in Turkey were evaluated for candidate genes at vernalization (*Vrn-A1*) and photoperiod (*Ppd-D1*) loci by using Competitive Allele Specific PCR (KASP) assays. We tried to unveil relative effects and interactions between genes for HD by determining the allelic distribution for *Vrn-A1* and *Ppd-D1* loci. This will help to improve breeding strategies by using genetic markers to develop new varieties suitable for changing climate conditions.

PE0890: Wheat, Barley, Oat, and related

Gains through Selection for Grain Yield in a Winter Wheat Breeding Program

Dennis Lozada, Washington State University, Pullman, WA, **Brian P. Ward**, USDA-ARS Plant Science Research, Raleigh, NC and **Arron H. Carter**, Dept. of Crop and Soil Science, Washington State University, Pullman, WA

Increased genetic gains for complex traits in plant breeding programs can be achieved through different selection strategies. The objective of this study was to compare potential gains for grain yield in a winter wheat breeding program through estimating response to selection R values across several selection approaches including phenotypic (PS), marker-based (MS), genomic (GS), and a combination of PS and GS. Five populations of Washington State University (WSU) winter wheat breeding lines evaluated from 2015 to 2018 for grain yield in Lind and Pullman, WA, USA were used in the study. Selection was conducted by selecting the top 20% of lines based on observed yield (PS strategy), genomic estimated breeding values (GS), presence of yield “enhancing” alleles of the most significant single nucleotide polymorphism (SNP) markers identified from genome-wide association mapping (MS), and high observed yield and estimated breeding values (PS+GS). Overall, PS compared to other individual selection strategies showed the highest mean response ($R = 0.61$). When combined with GS, a 23% improvement in R for yield was observed, indicating that gains could be improved by complementing traditional PS with GS. Negative responses were observed for MS and GS. MS was not successful in terms of R relative to the other selection approaches; using this strategy resulted in a significant ($P < 0.05$) decrease in response to selection compared to other methods. Gains through increased response to selection for yield could be achieved in the WSU winter wheat breeding program by implementing different selection strategies either exclusively or in combination.

PO0891: Wheat, Barley, Oat, and related

Characterization, Validation, and Deployment of the Chromosome 6BL QTL for Grain Yield Components in Hard Winter Wheat

Andrew M. Katz, Patrick Byrne, Scott Haley and Stephen Pearce, Colorado State University, Fort Collins, CO

The United Nations has estimated that food production will need to double by 2050 to adequately feed a global population of 9 billion people. Improvements in wheat yields, which account for 30% of coarse grain production, will be essential to meet this goal. Yield is a complex trait due to a multitude of influential factors. To address this complexity we have identified an individual yield component which is less complex and contributes to overall yield. A GWAS of a hard winter wheat association-mapping panel identified QTLs on the 7AL and 6BL chromosome arm for spikelet number per spike. The Great Plains winter wheat cultivar Platte and experimental line CO940610 were identified as polymorphic in the 7AL and 6BL regions. A population of recombinant inbred lines generated from the two parents, grown over the 2016 and 2018 seasons, validated the 7AL and 6BL QTLs' effect on spikelet number per spike. Confirmed SNPs were derived via exome sequencing data generated from the parental lines which will enable high-resolution mapping of the causative genetic variant underlying these QTL. These SNPs will be used to introgress spikelet number QTLs into two Colorado advanced lines and three high biomass lines from the International Maize and Wheat Improvement Center (CIMMYT). The employment of these novel genomic tools and resources enable unprecedented opportunities to identify allelic variation underlying individual yield components in wheat, and ultimately aiding in the development of higher yielding wheat varieties.

PE0892: Wheat, Barley, Oat, and related

High Density Genotyping and Diversity Analysis of a Historical Panel of Soft Winter Wheat Germplasm from the Eastern United States

Brian P. Ward¹, Alina Akhunova², Katherine Jordan³, Fei He⁴, Jie Ren⁵, Gina L Brown-Guedira¹ and Eduard Akhunov⁴, (1)USDA-ARS Plant Science Research, Raleigh, NC, (2)Integrated Genomics Facility, Kansas State University, Manhattan, KS, (3)Kansas State University, Department of Plant Pathology, Manhattan, KS, (4)Department of Plant Pathology, Kansas State University, Manhattan, KS, (5)Integrated Genomics Facility, Manhattan, KS

The vast majority of wheat (*Triticum aestivum* L.) grown in the Eastern United States falls into the market class of soft red winter wheat. While the recent publication of the Chinese Spring reference genome has enabled a wide range of genomics-based analyses, soft red winter wheats remain underrepresented in publicly-available data repositories. Although total Eastern U.S. wheat acreages are lower than that of the hard winter and spring wheats grown in the Great Plains, this region represents a highly diverse set of germplasm adapted to a unique set of biotic and abiotic stresses. Traits that are important to the success of soft red winter wheat include resistance to high fungal disease pressure, resistance to several herbivorous insect species including Hessian fly (*Mayetiola destructor* Say), and escape from high temperatures in late spring and early summer via shortened vernalization requirement and photoperiod insensitivity. Favorable alleles for these traits may currently be absent from many publicly-available genomic data repositories. Previously, exome capture sequencing was performed on the soft winter wheat cultivars AGS-2000, MPV57, LA95135, and Pioneer Brand 26R61, all of which have served as common parents in Eastern United States breeding programs. In order to identify candidates for additional exome capture sequencing to represent a broad portion of soft red winter wheat diversity, we analyzed genotyping-by-sequencing (GBS) data on an additional 2,117 accessions, representing landraces, historically important cultivars, and recently released lines which are commonly used as parents in current regional breeding programs. After examining the available data and soliciting input from Eastern U.S. wheat breeders, we selected 54 lines for additional exome capture. In addition, we reran GBS on these lines to create matching sets of exome capture and GBS SNPs derived from the same tissue samples. These lines, together with the four previously-sequenced lines, capture a large portion of the genetic diversity present in Eastern U.S. breeding programs. We demonstrate how this new diversity panel can be used to mine coding sequence variation present in soft winter wheat germplasm, identify ancestral alleles and haplotypes, and impute high density marker data onto lower density data generated by arrays, targeted amplicon sequencing, or GBS.

PO0893: Wheat, Barley, Oat, and related

Adaptation of Winter Wheat in the Southeastern U.S. is Associated with Unique Flowering Time Alleles

Gina L Brown-Guedira¹, Mohammed Guedira², Mai Xiong³, Jose Martin Sarinelli³, Noah DeWitt³ and Brian P. Ward¹, (1)USDA-ARS Plant Science Research, Raleigh, NC, (2)Department of Crop & Soil Science, Raleigh, NC, (3)North Carolina State University, Raleigh, NC

Flowering time is an important trait in wheat breeding as it affects adaptation and yield potential. In the southeastern USA, where winters are characterized by variable periods of warm weather that may be followed by cold, wheat cultivars that flower too early have an increased risk of frost damage. In contrast, cultivars that flower too late have increased risk of high temperature stress and water deficit, which can restrict grain formation and consequently reduce yield. Multiple mapping studies involving early flowering eastern soft red winter (SRW) wheat cultivars have identified QTL for heading date associated with variation at the *VRN1* and *PPD1* loci. These include a winter allele of *Vrn-B1* having a 36 bp intron one deletion associated with early flowering after shorter periods of cold treatment. Evaluation of diverse winter wheat germplasm determined this allele was rare in global germplasm but was detected in the North American landrace Purplestraw and is common in modern southeastern wheat cultivars. In addition, Purplestraw possessed the globally rare *Ppd-A1a.1* allele having a 1,085 bp deletion in the critical upstream region that is present in approximately 40% of contemporary eastern wheats. Evaluation of genotypic and phenotypic data from multiple years of collaborative SRW wheat yield trials revealed winter alleles of *Vrn-A1* and *Vrn-B1* requiring shorter periods of cold to potentiate flowering in combination with insensitive *PPD1* alleles provide a grain yield advantage at southern locations. In the Mid-Atlantic region, longer vernalization alleles were desirable. Our results demonstrate that *VRN1* and *PPD1* alleles of varying strength deployed in combination with genes of smaller effect allow fine tuning of flowering time in diverse winter wheat growing environments.

PE0894: Wheat, Barley, Oat, and related

QTL Mapping of Forage Yield Traits in Winter Wheat

Frank Maulana¹, Joshua D. Anderson¹, Wangqi Huang^{1,2}, Habtamu Ayalew Tamir¹, Tadele Kumssa¹, Shuyu Liu³ and Xuefeng Ma¹, (1)Noble Research Institute, LLC, Ardmore, OK, (2)State Key Laboratory for Conservation and Utilization of Bio-Resources in Yunnan, Yunnan Agricultural University, Kunming, China, (3)Texas A&M AgriLife Research, Amarillo, TX

Wheat is one of the most important cereal crops grown in the Southern Great Plains of the United States. In this region, winter wheat is often grown for dual-purpose for grazing animals during the autumn-winter seasons and grain production by summer. Phenotyping forage yield traits is time-consuming and expensive. Marker-assisted selection (MAS) may provide a viable and more efficient selection approach to facilitate breeding selection of the traits. A QTL mapping study was conducted to map QTL and identify molecular markers associated with forage traits in winter wheat. A recombinant inbred line (RIL) mapping population, CT111 (n = 214), developed from a cross between TAM 111 and CO960293-2, was used in this study. The population was genotyped with 4990 SNP markers and phenotyped for various agronomic traits related to forage yield in four environments. With inclusive composite interval mapping (ICIM) analysis performed using QTL IciMapping v4.1, significant QTL were detected on different chromosomes. The molecular markers that are tightly linked to the identified QTL will facilitate candidate gene discovery and marker-assisted selection in dual-purpose wheat breeding.

PO0895: Wheat, Barley, Oat, and related

Optimizing Training Population Composition to Improve Prediction Accuracy in Wheat

Emmanuel Adeyemo, Emily J. Conley, Rex Bernardo and James A Anderson, University of Minnesota, St. Paul, MN

Maximizing the relationship between the training and breeding populations have shown to improve the accuracy of genomic prediction. In this study, four different sampling algorithms; stratified sampling, genomic relationship, mean of coefficient of determination (CDmean) and random sampling were evaluated for prediction accuracies. To accomplish this, 383 University of Minnesota wheat breeding lines tested across seven locations in Minnesota over a period of five years, were used to train the four different models. These lines were evaluated for yield, test weight, protein content and plant height. In this poster, we will show detailed results on the effectiveness of each sampling algorithm for training population selection.

PE0896: Wheat, Barley, Oat, and related

GWAS and Hyperspectral Reflectance: Association Mapping of Photosynthetic/Photoprotective Pigment Composition in Spring Wheat

Ryan Joynson¹, Gemma Molero², Laura-Jayne Gardiner³, Carolina Rivera-Amado², John Evans⁴, Robert Furbank⁴, Matthew Reynolds² and Anthony Hall¹, (1)Earlham Institute, Norwich, United Kingdom, (2)International Maize and Wheat Improvement Center (CIMMYT), Texcoco, Mexico, (3)IBM Research, Warrington, United Kingdom, (4)Australian National University, Canberra, Australia

One of the major challenges plant scientists face is increasing wheat (*Triticum aestivum*) yield potential (YP). A significant bottleneck for increasing YP is achieving increased biomass through optimization of radiation use efficiency (RUE) along the crop cycle. The composition and ratios of both photosynthetic and photoprotective pigments in leaves can vary greatly in wheat and are thought to have a significant effect on photosynthetic efficiency; however, the effect of this diversity on YP is still unclear. To understand the genetic basis of leaf pigment composition and its links to RUE, we have conducted a genome-wide association study (GWAS) using a panel of 150 elite spring wheat genotypes including landrace and synthetically derived lines assembled at CIMMYT. The panel was evaluated for yield, biomass and RUE related traits alongside pigmentation analysis using hyperspectral reflectance. Phenotyping was carried out over 2 years under optimal growing conditions at the International Wheat Yield Partnership (IWYP) hub in Obregon, Mexico. Leaf reflectance curves were used to quantify more than 30 traits relating to pigmentation, carbohydrate content and water content. The panel was genotyped using a custom 12Mb hybridisation enrichment capture which yielded an average of 700K SNPs per accession. Marker-trait association identified multiple genomic regions significantly associated with leaf pigment levels. This included photosynthetic pigments such as chlorophyll A and B and also photoprotective pigments such as carotenoids. Allelic variation in SNPs associated with pigmentation variation in a number of traits could be attributed to the presence of exotic material in the pedigree history of the panel members. This demonstrates the value of integrating high RUE exotic material into global prebreeding programs as a strategy to optimise RUE in elite varieties.

PO0897: Wheat, Barley, Oat, and related

Multi-Cultivar Wheat Tetraploid and Hexaploid Exome Capture with a Single Enrichment Kit

Jacob Enk, Lauren O'Connor, **Alison Devault** and Jean-Marie Rouillard, Arbor Biosciences, Ann Arbor, MI

Hybridization capture technology for targeted sequencing is highly tolerant of sequence variation between probes and targets. The user can tune experimental conditions to balance specificity and sensitivity, so that probes designed from one taxon can be used to enrich DNA from closely related taxa with high efficiency. Here we demonstrate how the Arbor Biosciences myBaits Wheat Exome panel, developed with IWGSC using the hexaploid Chinese Spring bread wheat reference genome, successfully enriches exomes of a broad range of wheat cultivars, including tetraploid varieties. We anticipate that large portions of the exomes of other cereal crops, including rye and barley, can also be obtained with this wheat-derived probe set under even default experimental conditions. This translates to inexpensive and logistically simple exome sequencing for projects of any cultivar composition, stage or scale.

PE0898: Wheat, Barley, Oat, and related

Wheat Doubled Haploids have a Marked Prevalence of Deletions and Aneuploidy

Sandesh Kumar Shrestha¹, Curtis Pozniak², Pierre J. Hucl², Sean Walkowiak³, Byron Evers¹, Shuangye Wu¹ and Jesse Poland¹, (1)Department of Plant Pathology, Kansas State University, Manhattan, KS, (2)Crop Development Centre, University of Saskatchewan, Saskatoon, SK, Canada, (3)Grain Research Laboratory, Canadian Grain Commission, Canada

Understanding the structural variation such as deletions in crop varieties is crucial as it often alters phenotypic expression of agronomic importance. Likewise, aneuploidy is genetically unstable and causes severe fitness penalties due to compromised gene dosage. Wheat doubled haploids (DH) are rapidly fixed valuable breeding materials that can be quickly evaluated and returned as parents. However, DH line development procedure, generation of haploid plant and chromosome doubling, can induce genomic changes. To address the underlying

hypothesis of impact of such structural changes, we studied a DH population developed by wheat-maize wide hybridization derived from intercrossing of Canadian spring wheat cultivars CDC Stanley and CDC Landmark, both of which have a complete assembled genome. We first used MUMMer to identify structural variations between two reference genomes and verified them using read depth generated by skim-sequencing of 1x coverage for each of 48 DH lines along with high coverage sequencing of the two parents. After read alignment to CDC Landmark reference genome, the number of reads assigned to 100Kb bins were normalized to 1x coverage based on sample coverage. Any deviation from the expected 1x coverage to 0x in a specific region or 0.5x along the complete chromosome shows deletion or only a single chromosome copy (aneuploidy), respectively. The correlation of results obtained by genome alignment and read depth along with aneuploidy observed in DH lines revealed extensive deletions on most chromosomes in the primary DH lines, and unexpected instances of aneuploidy. The impact of these changes on breeding will be discussed.

PO0899: Wheat, Barley, Oat, and related

The Bread Wheat Epigenomic Map Reveals Distinct Chromatin Architectural and Evolutionary Features of Functional Genetic Elements

Yi-Jing Zhang, Institute of Plant Physiology and Ecology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai, China

Bread wheat is an allohexaploid species with a 16-Gb genome that has large intergenic regions, which presents a big challenge for pinpointing regulatory elements and further revealing the transcriptional regulatory mechanisms. Chromatin profiling to characterize the combinatorial patterns of chromatin signatures is a powerful means to detect functional elements and clarify regulatory activities in human studies.

Here, through comprehensive analyses of the open chromatin, DNA methylome, seven major chromatin marks, and transcriptomic data generated for seedlings of allohexaploid wheat, we detected distinct chromatin architectural features surrounding various functional elements, including genes, promoters, enhancer-like elements, and transposons. Thousands of new genic regions and cis-regulatory elements are identified based on the combinatorial pattern of chromatin features. Roughly 1.5% of the genome encodes a subset of active regulatory elements, including promoters and enhancer-like elements, which are characterized by a high degree of chromatin openness and histone acetylation, an abundance of CpG islands, and low DNA methylation levels. A comparison across sub-genomes reveals that evolutionary selection on gene regulation is targeted at the sequence and chromatin feature levels. The divergent enrichment of cis-elements between enhancer-like sequences and promoters implies these functional elements are targeted by different transcription factors. Altogether, we present a systematic epigenomic map for the annotation of cis-regulatory elements in the bread wheat genome, which provides new insights into the connections between chromatin modifications and cis-regulatory activities in allohexaploid wheat.

All data were visualized with a customized genome browser (http://bioinfo.sibs.ac.cn/cs_epigenome).

PE0900: Wheat, Barley, Oat, and related

Imputation and Mapping in Multiparental Crop Populations from Low Coverage Sequence Data

Michael F Scott, UCL, London, United Kingdom

We report genetic variation segregating in Multi-parent Advanced Generation Inter-Cross (MAGIC) populations of wheat and rice. The wheat population was derived from sixteen diverse wheat varieties released in Northern Europe between 1935 and 2004, which were sequenced after genic and promoter capture enrichment to give 20x coverage of the target regions. The rice heat MAGIC population is derived from eight founder varieties that were selected on the basis of temperature tolerance and sequenced to 13.5x. We use these data to identify >1m polymorphic sites and then imputed the genotypes of Recombinant Inbred Lines (RILs), which were sequenced at low coverage: 504 wheat RILs sequenced at 0.3x coverage and 836 rice RILs sequenced at 1.4x coverage. We achieved a high call rate (99.5%) and accuracy (99.1%) through imputation. Downsampling indicates that the call rate and genotyping accuracy after imputation would have been comparably high with coverage as low as 0.1x. We use these data to identify large introgressions and fine-map QTLs for traits of agronomic importance. Our results suggest that

haplotype-based imputation from low coverage whole genome sequence data, with or without the use of a reference haplotype panel, is an effective genotyping strategy in crop populations.

PO0901: Wheat, Barley, Oat, and related

Genome Diversity of Central European Wheat Landrace Collection Compared to Elite Bread Wheat Cultivars

András Cseh¹, Péter Poczai², Tibor Kiss³, Krisztina Balla¹, Zita Berki¹, Ádám Horváth¹ and Ildikó Karsai³, (1)Agricultural Institute, Centre for Agricultural Research, Martonvásár, Hungary, (2)Organismal Evolution and Biology, Viikki Plant Science Centre, University of Helsinki, Helsinki, Finland, (3)Centre for Agricultural Research, Agricultural Institute, Martonvásár, Hungary

Wheat landraces (*Triticum aestivum* L.) represent a valuable basis of genetic diversity and specific adaptation to local environmental conditions according to their place of origin. Their population genetic structure, buffering capacity, and a combination of morpho-physiological traits implies a better tolerance to changing climate conditions and to stress environments than modern cultivars. A large set of Central European landraces were collected during the 1950-60s and preserved in the Hungarian Gene Bank collection (Tápiószele). The collection represents an important legacy and provides an untapped source of genetic variation for wheat improvement.

The aim of the present work is to provide detailed genotyping data for identification of the valuable source of genetic diversity on the Central European landrace collection encompassing 200 landraces in parallel to 70 modern wheat cultivars. Genotyping was carried out using a high-density Illumina SNP genotyping array with 17,905 gene-based SNP probes. We analysed the geographic distribution of the genetic variability and chromosome distribution of polymorphic markers. The landrace accessions had a greater percentage of polymorphic markers on the homoeologous group 2 and 7 chromosomes. The group 2 and 7 chromosomes contain several genes of great agronomic importance including the photoperiod response genes, semi-dwarfing genes, fungal disease resistance genes and the thousand-grain weight gene. Structure and principal component analysis showed considerable differences among landraces and modern wheat cultivars, thus the Central European wheat landrace collection could be a valuable reservoir of genetic variability suitable for future breeding programmes.

AC gratefully acknowledges a Marie Skłodowska-Curie Fellowship Grant (H2020-MSCA-IF-2016-752453-LANDRACES) from the European Union. This study was financially supported by National Research, Development and Innovation Office of Hungary (K129221).

PE0902: Wheat, Barley, Oat, and related

Do Spatial Designs Outperform Classic Experimental Designs?

Raegan Hoefler¹, Pablo M Gonzalez Barrios¹, Madhav Bhatta¹, Inés Berro¹, Rafael Storto Nalin², Alejandra Borges³, Eduardo Covarrubias⁴, Luis Diaz Garcia⁵ and Lucia Gutierrez¹, (1)Department of Agronomy, University of Wisconsin - Madison, Madison, WI, (2)Department of Genetics, University of Sao Paulo - "Luiz de Queiroz" College of Agriculture, Piracicaba, Brazil, (3)Statistics Department, Facultad de Agronomía, Universidad de la República, Montevideo, Uruguay, (4)CGIAR Excellence in Breeding Platform (EiB)/International Maize and Wheat Improvement Center (CIMMYT), El Batán, Mexico, (5)Instituto Nacional de Investigaciones Forestales, Agrícolas y Pecuarias, Aguascalientes, Mexico

Controlling spatial variability in agricultural field trials is the most important step to efficiently and accurately compare genotypes. Spatial variability can be controlled at the experimental design level with the assignment of genotypes to experimental units and at the modeling level with the use of spatial corrections, the relationship among treatments, and other modeling strategies. The goal of this study was to use a simulation approach with true wheat performance data to compare the efficiency of complex randomization-based experimental designs with the inclusion of various spatial correction methods and the correlation between genotypes under several field scenarios. We evaluated the effect of plot size, experiment size, genotype by environment interaction structure, and heritability on the performance of eleven experimental designs with each of four spatial correlation structures. Spatial

experimental designs outperformed the classic experiments in terms of predictive ability but not in terms of response to selection, for which partially replicated experiments with larger population sizes were superior under most scenarios. An AR1 was the superior spatial correction model for some of the experimental designs in the medium-sized experiments, but no differences in performance of any spatial correction models were found for the small experiments. Our results provide a clear guide to researchers designing and analyzing large field experiments.

PO0903: Wheat, Barley, Oat, and related

Increased Prediction Accuracy Using Combined Genomic Information and Physiological Traits in a Soft Wheat Panel Evaluated in Multi-Environments

Jia Guo¹, Sumit Pradhan Shrestha¹, Dipendra Shahi¹, Jahangir Khan¹, Jordan McBreen², Guihua Bai³, Paul Murphy⁴ and Md A. Babar¹, (1)University of Florida, Gainesville, FL, (2)University of Florida, GAINESVILLE, FL, (3)USDA, Hard Winter Wheat Genetics Research Unit, Manhattan, KS, (4)North Carolina State University, Raleigh, NC

An integration of field-based phenotypic and genomic data can potentially increase the genetic gain in wheat breeding for complex traits such as grain and biomass yield. To validate this hypothesis in empirical field experiments, we compared the prediction accuracy between multi-kernel physiological and genomic model to a single-kernel physiological or genomic model for grain yield (GY) using a soft wheat population that was evaluated in four environments. The physiological data including canopy temperature (CT), SPAD chlorophyll content (SPAD), membrane thermostability (MT), rate of senescence (RS), stay green trait (SGT), and NDVI values were collected at four environments (2016, 2017, and 2018 at Citra, FL; 2017 at Quincy, FL). Using a genotyping-by-sequencing (GBS) approach, a total of 19,353 SNPs were generated and used to estimate prediction model accuracy. Prediction accuracies of grain yield evaluated in four environments improved when physiological traits and/or interaction effects (genotype \times environment or physiology \times environment) were included in the model compared to models with only genomic data. The proposed multi-kernel models that combined physiological and genomic data showed 35 to 169% increase in prediction accuracy compared to models with only genomic data included when heading date was used as a covariate. The results of this study support the integration of field-based physiological data into GY prediction to improve genetic gain from selection in soft wheat under a multi-environment context.

PE0904: Wheat, Barley, Oat, and related

Can We Use Machine Learning to Predict Circadian Genes in Wheat using DNA Sequence?

Laura-Jayne Gardiner¹, Hannah Rees², Rachel Rusholme-Pilcher², Susan Duncan², Anna-Paola Carrieri¹, Anthony Hall² and Ritesh Krishna³, (1)IBM Research, Warrington, United Kingdom, (2)Earlham Institute, Norwich, United Kingdom, (3)International Business Machines, Manchester, United Kingdom

We focus on the biological process of circadian regulation that has been found to underpin many agronomic traits in wheat, a key crop of global importance. Genes involved in the circadian clock show rhythmic expression patterns of approximately 24 hours that can be defined using parameters such as period, phase and amplitude. We use a 48-hour time course transcriptomics dataset generated by the Earlham Institute to identify 30,065 high confidence genes that are likely to be circadian in wheat.

We demonstrate the use of machine learning approaches for classification of the time-series expression profiles of genes into one of five classes [morning-circadian, day-circadian, evening-circadian, night-circadian or non-circadian] with an average accuracy of 85%. Furthermore, our accuracy is maintained (80%) using only 12 of the 24 timepoints available, where other commonly used tools showed an accuracy of only 64.5%. Now we report our exploration of the possibility of assigning genes into our five classes based on DNA sequence using features such as enriched putative regulatory DNA elements or motifs, SNPs or epigenetic marks.

This methodology can be applied to a wide range of genomics problems to reduce the time, cost and effort to identify patterns using predictive models. This work was supported by the STFC Hartree Centre's Innovation Return on Research programme, funded by the UK Department for Business, Energy & Industrial Strategy and is part of an

ongoing collaboration between IBM Research, UK, and the Earlham Institute. IBM brings cutting-edge computational science alongside applicable technologies to support UK research.

PO0905: Wheat, Barley, Oat, and related

Characterisation of Functional lncRNA in *T. aestivum* Strains

ThankGod E. Ebenezer, Benjamin H White, Susan Duncan, Anthony Hall and Wilfried Haerty, Earlham Institute, Norwich, United Kingdom

Common bread wheat (*Triticum aestivum*) has evolved through two separate allopolyploidy events ~500,000 and ~10,000 years ago resulting in an hexaploidy genome. These sub-genomes contain complete, yet evolutionarily divergent, sets of duplicated coding and non-coding genes. These non-coding genes form a widely heterogeneous class of transcripts and are generally involved in gene expression regulation. While we have a relatively good understanding of the biological role of short (<200nt) RNA, much remains to be learned for long (> 200 nt) noncoding RNA (lncRNA). Most of the studies focusing on animal genomes and few model plant organisms (e.g. Arabidopsis, rice) highlighted the rapid evolutionary turnover of these genes as well as the functional role of few of them. Here, we investigate *Triticeae* species with different ploidy levels to assess the evolution, birth and death of lncRNAs with respect to ploidy levels and hybridizations event. Additionally, we are leveraging on newly generated transcriptomics datasets across elite wheat varieties to characterise lncRNAs variation within this species.

PE0906: Wheat, Barley, Oat, and related

Prediction of Wheat-Rye 1RS Translocations (1AL.1RS, 1BL.1RS) and the Impact on Wheat Breeding

Liangliang Gao¹, M. Timothy Rabanus-Wallace², Martin Mascher², Nils Stein², Allan K. Fritz³ and Jesse Poland¹, (1)Department of Plant Pathology, Kansas State University, Manhattan, KS, (2)Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Gatersleben, Germany, (3)Department of Agronomy, Kansas State University, Manhattan, KS

The whole arm (Robertsonian) translocations of rye 1RS to wheat chromosomes 1A and 1B are known to possess various biotic and abiotic resistance traits, yet with reduced bread making quality particularly for the translocation on wheat 1B. The availability of a rye reference genome enabled us to examine the group 1 wheat-rye translocations in detail. We first delineated the wheat-rye 1BL.1RS_{Kavkaz} and 1AL.1RS_{Amigo} translocation segment to be approximately 269Mb, directly at the projected centromere positions, confirming whole chromosome arm translocations. From this we developed a bioinformatics pipeline to predict the presence or absence of the wheat-rye translocations based on GBS or exome sequencing data for various wheat panels globally. We detected 1BL.1RS in varying frequencies (6.5%-31%) for central US winter wheat panels (>4000 lines), European WHEALBI panel (>500 lines) and CIMMYT spring wheat breeding panel (>900 lines). In contrast, 1AL.1RS translocation is only detected at 4%-10% in central US materials but not in European or CIMMYT panels. We called SNPs and calculated the identity by state percentages among various 1RS lines. Our results suggest that the two 1R translocations found in thousands of breeding lines in different panels globally likely share a single common origin designed at 1RS_{Kavkaz} and 1RS_{Amigo}. We found positive correlations for 1RS and grain yield for central US winter wheat materials. Importantly, we identified a novel 1R recombination line between 1BL.1RS_{Kavkaz} and 1AL.1RS_{Amigo} that shows high yield and is not associated with very poor bread making quality. The 1RS prediction pipeline developed will enable breeding programs to monitor the presence of rye translocations. Our work also demonstrates the potential of targeted breeding of 1RS and other rye translocations to advance wheat breeding for productivity, resilience and good quality.

PO0907: Wheat, Barley, Oat, and related

Development and Application of Genome-Specific SNP Markers for Tracing Alien Introgressions in Polyploid Wheat Genome

Tatiana V. Danilova¹, Wei Zhang², Mingyi Zhang², Xianwen Zhu², Jesse Poland³, Jason D. Fiedler⁴, Bernd Friebe⁵ and Xiwen Cai², (1)North Dakota State University, Fargo, ND, (2)Department of Plant

Science, North Dakota State University, Fargo, ND, (3)Department of Plant Pathology, Kansas State University, Manhattan, KS, (4)USDA-ARS, Cereal Crops Research Unit, Edward T. Schafer Agricultural Research Center, Fargo, ND 58102, Fargo, ND, (5)Kansas State University, Manhattan, KS

New traits can be introduced into crops through interspecific hybridization. This approach has been successfully applied for wheat improvement. Detection of wheat-alien introgressions requires screening of large populations and is time and labor consuming. With next generation sequence resources available for wheat and related species, single nucleotide polymorphism (SNP) markers provide an effective tool for detecting alien introgressions. The allopolyploidy of the wheat genome ($2n=6x=42$, AABBDD) makes introgressions and chromosome manipulations possible, but complicates the development of genome-specific co-dominant molecular markers. We found that the four-genome-specific allelic SNPs needed for developing molecular markers are rare, whereas closely located two-genome-specific SNPs are more common. These “shifted” SNPs do not need much sequence data to discover and can be used for developing genotyping assays. Chromosomal locations of sequences containing SNPs are important for tracing recombination events by molecular markers. The wheat cDNA cytogenetic map is a useful resource for developing molecular markers with known positions. Mapped cDNAs cover all chromosomes of the three wheat sub-genomes, and orthologous sequences can be found in sequenced genomes of related species. PCR Allelic Competitive Extension (PACE) genotyping assays with co-dominant shifted SNP markers were developed using mapped sequences and applied to trace barley, *Aegilops speltoides*, *Thinopyrum elongatum* and *Th. intermedium* introgressions in hexaploid wheat background. This approach improved the throughput and accuracy in detecting homoeologous recombinants and tracing alien introgressions in wheat.

PE0908: Wheat, Barley, Oat, and related

Dissection and Cytological Mapping of an *Ae. speltoides*-Originated Gene for Stunted Growth in Wheat

Wei Zhang¹, Mingyi Zhang¹, Xianwen Zhu², Shuangfeng Ren², Jason D. Fiedler³, Steven S. Xu⁴ and Xiwen Cai¹, (1)Department of Plant Science, North Dakota State University, Fargo, ND, (2)North Dakota State University, Fargo, ND, (3)USDA-ARS, Cereal Crops Research Unit, Edward T. Schafer Agricultural Research Center, Fargo, ND 58102, Fargo, ND, (4)USDA-ARS, Fargo, ND

Aegilops speltoides ($2n=2x=14$, SS) has been considered one of the possible donors for wheat B genome. It has contributed to wheat origin and evolution as well as favorable genes to wheat improvement. In this study, we identified a gene on *Ae. speltoides* chromosome 2S that resulted in stunted growth of wheat (*Triticum aestivum* L., $2n=6x=42$, AABBDD). Chromosome 2S was engineered by inducing homoeologous recombination with wheat chromosome 2B. Totally 98 2B-2S recombinants were produced, phenotyped for plant growth, and delineated by GISH and SNP assay. The stunted growth gene physically mapped to the sub-telomeric region (~ 5 Mb) on the short arm of chromosome 2S and designated *SgAes1*. The 2B-2S recombinants containing *SgAes1* showed stunted growth, while those without *SgAes1* showed normal growth. We transferred the recombinant chromosome 2SS-2BS-2SS-2SL, which had the 2S segment containing *SgAes1* replaced by its homoeologue of chromosome 2B, to *Ae. speltoides* by marker-assisted backcrossing. The plants with the native chromosome 2S substituted by 2SS-2BS-2SS-2SL exhibited similar growth and morphology as wild type *Ae. speltoides*. Thus, the substitution of *SgAes1* for its homoeologue on chromosome 2B in the wheat background led to stunted growth. But the replacement of *SgAes1* by its wheat homoeologue in the *Ae. speltoides* background did not lead to obvious change in plant growth. Apparently, *SgAes1* underwent evolutionary modifications in wheat if the wheat homoeologue of *SgAes1* was derived from *Ae. speltoides*. Further studies are underway to better understand the function of *SgAes1* and its implication in wheat origin and evolution.

PO0909: Wheat, Barley, Oat, and related

Locating Alien Introgression in Wheat Using Long and Short Read Technologies

Benedict Coombes¹, Ryan Joynson¹, Julie King², Surbhi Grewal³, Stella Hubbard³, John Fellers⁴, Ian P. King³ and Anthony Hall¹, (1)Earlham Institute, Norwich, United Kingdom, (2)University of Nottingham, Loughborough, Leicestershire, United Kingdom, (3)The University of Nottingham, Loughborough, United Kingdom, (4)USDA ARS, Manhattan, KS

Wild relatives of wheat possess a broad pool of variation which could be utilised to overcome the historic genetic bottlenecks in current global elite wheat populations. The wild relatives have been observed to have many favourable agronomic traits, particularly relating to tolerance to biotic and abiotic stresses where they often surpass elite varieties. Despite its high potential utility to breeders around the world, this variation has largely been underutilised to date. Collaborators at the University of Nottingham have produced a set of double haploid wheat lines containing introgressions from the wheat wild relative *Ambylopyrum muticum*. However, in order to study these introgressions and begin to utilize them in breeding programmes, we first must be able to reliably locate them to a high resolution. Marker analysis with the 35K Axiom® Wheat-Relative Genotyping Array has previously been used to detect introgressions in these lines but lacks the resolution to precisely locate the introgression borders and may miss small introgressions entirely. To facilitate high resolution detection of introgressions we have utilized short read whole genome sequencing data from Am. Muticum and Oxford Nanopore data from the introgression wheat lines. Here we describe our current progress, integrating both long and short read technologies to identify introgression borders and their genic content.

PE0910: Wheat, Barley, Oat, and related

Introgression and Expression of Tall Wheatgrass (*Thinopyrum ponticum*) Genes in the Polyploid Genome of Hexaploid Wheat (*Triticum aestivum*)

Kirby Nilsen¹, Sean Walkowiak², **Valentyna Klymiuk**¹, Martha Bonilla Betancourt¹, Peter Langridge³, Ken Chalmers⁴, Pierre J. Hucl², Richard Cuthbert⁵ and Curtis Pozniak², (1)Department of Plant Sciences, University of Saskatchewan, Saskatoon, SK, Canada, (2)Crop Development Centre, University of Saskatchewan, Saskatoon, SK, Canada, (3)The University of Adelaide, Adelaide, Australia, (4)University of Adelaide, Adelaide, Australia, (5)Swift Current Research and Development Centre, Agriculture and Agri-Food Canada, Swift Current, SK, Canada

Bread wheat (*Triticum aestivum*) and durum wheat (*T. durum*) underwent a severe genetic bottleneck during domestication and breeding and their secondary and tertiary gene pools are a rich reservoir of genetic diversity that can be used for crop improvement. As part of the 10+ Wheat Genome Project, we have generated and made available whole genome assemblies for 16 wheat genotypes (www.10wheatgenomes.com). A detailed comparison of these assemblies identified an approx. 60 Mb terminal introgression on the long arm of chromosome 3D in the elite Australian cultivar Lancer. Based on pedigree analysis, we hypothesized the introgressed chromatin may have derived from tall wheatgrass (*Thinopyrum ponticum*), the source of major leaf (*Lr24*) and stem rust resistance genes (*Sr24*). We performed short read (Illumina paired-end) and/or long read (Oxford nanopore) whole genome shotgun sequencing of *T. ponticum* cv. Orbit and two additional wheat lines thought to carry the introgression, AAC Concord and GP091. Mapping of the sequencing reads to the Lancer assembly confirmed that the introgression in Lancer is derived from *T. ponticum* and the introgressions in AAC Concord and GP091 are identical by descent. Furthermore, we performed RNA sequencing of eight independent wheat genetic backgrounds and comparative gene expression analysis identified *T. ponticum* derived genes whose expression may be impacting agronomic performance and resistance to abiotic and biotic stresses in wheat.

PO0911: Wheat, Barley, Oat, and related

Partitioning and Physical Mapping of Wheat Chromosome 3B and Its Homoeologue 3E in *Thinopyrum Elongatum* By Inducing Homoeologous Recombination

Mingyi Zhang¹, Wei Zhang¹, Xianwen Zhu¹, Qing Sun², Shiaoan Chao³, Changhui Yan², Jason D. Fiedler³, Steven S. Xu³ and Xiwen Cai¹, (1)Department of Plant Science, North Dakota State University, Fargo, ND58108, (2)Department of Computer Science, North Dakota State University, Fargo, ND58108, (3)USDA-ARS, Cereal Crops Research Unit, Edward T. Schafer Agricultural Research Center, Fargo, ND 58102

Thinopyrum elongatum (2n=2x=14, EE) genome is homoeologous to the wheat genome (AABBDD), and can be differentiated from wheat subgenomes by genomic *in situ* hybridization (GISH) and molecular markers. This allows one to partition the genomes by inducing meiotic homoeologous recombination for genome mapping and alien introgression. In this study, we constructed a special wheat genotype, which was double monosomic for wheat chromosome 3B and *Th. elongatum* chromosome 3E and homozygous for *ph1b* mutant, to induce 3B-3E

homoeologous recombination. Totally, 81 3B-3E recombinants involving different chromosome regions were recovered and detected in the primary, secondary, and tertiary homoeologous recombination by GISH. Comparing to the primary recombination, the secondary and tertiary recombination shifted toward the proximal regions due to the increase of homology between pairing partners. All recombinants were genotyped by high-throughput wheat 90K SNP arrays and recombination breakpoints were physically mapped based on GISH patterns and SNP genotyping results of the recombinants. As a result, 3B-3E recombination physically partitioned chromosome 3B into 38 bins, and 429 SNPs were assigned to the distinct bins. Integrative analysis of GISH and SNP genotyping results led to the construction of a composite bin map for chromosome 3B. Additionally, we developed 22 SNP-derived PCR markers specific for different regions of chromosome 3E and constructed a comparative map of chromosomes 3E, 3B, 3A, and 3D. Therefore, this work provides a unique physical framework to this homoeologous pair for further studies. Also, these homoeologous recombinants diversify and enrich the wheat genome for wheat improvement.

PE0912: Wheat, Barley, Oat, and related

Resistance to Wheat Rusts Identified in Wheat/*Ambylopyrum muticum* Chromosome Introgressions

John Fellers¹, Angie Matthews², Allan K. Fritz³, Matthew N. Rouse⁴, Surbhi Grewal⁵, Stella Hubbard⁵, Ian P. King⁵ and Julie King⁶, (1)USDA ARS, Manhattan, KS, (2)Kansas State University, Manhattan, KS, (3)Department of Agronomy, Kansas State University, Manhattan, KS, (4)USDA-ARS Cereal Disease Lab, St Paul, MN, (5)The University of Nottingham, Loughborough, United Kingdom, (6)University of Nottingham, Loughborough, Leicestershire, United Kingdom

Wheat rusts are a worldwide production problem. Plant breeders have used genetic resistance to combat these fungi. However, single-gene resistance is rapidly overcome due to the frequent occurrence of new virulent fungal strains so that a supply of new resistance sources is needed. New resistance sources are also limited within hexaploid genetic stocks. Wild relatives provide a new source of resistance genes. Twenty-eight hexaploid wheat/*Ambylopyrum muticum* introgression lines, with introgressions covering the majority of the T genome, were evaluated for resistance to *Puccinia triticina*, *P. graminis* f. sp. *tritici*, and *P. striiformis* f. sp. *tritici*. At the seedling level, four lines were resistant to races of *P. triticina*, six lines were resistant to *P. graminis*, and fifteen lines were resistant to *P. striiformis*. At the adult stage, sixteen lines were resistant to *P. triticina*. Some of these lines will require further work to reduce the size of the introgressed segment, however, lines 92 and 355 have very small fragments and can be used directly as new resistance donors.

PO0913: Wheat, Barley, Oat, and related

Dissecting the Sea Wheatgrass Genome to Transfer Biotic Stress Resistance and Abiotic Stress Tolerance into Wheat

Dilkaran S. Singh¹, Qijun Zhang², Marie Langham¹, Steven S. Xu³ and **Wanlong Li¹**, (1)South Dakota State University, Brookings, SD, (2)North Dakota State University, Fargo, ND, (3)USDA-ARS, Fargo, ND

Wheat production is facing numerous challenges from biotic and abiotic stresses. Alien gene transfer has been an effective approach for wheat germplasm enhancement. Sea wheatgrass (SWG) (*Thinopyrum junceiforme*, 2n = 4x = 28, genomes J₁J₁J₂J₂), is a distant relative of wheat and a relatively untapped source for wheat improvement. We have identified high tolerance to waterlogging, manganese toxicity, heat and low nitrogen and resistance to wheat streak mosaic virus (WSMV), Fusarium head blight and wheat stem sawflies (due to the solid stem) in SWG. Our **long-term goal** is to broaden the wheat genetic basis and develop SWG-derived novel germplasm that will contribute to a more sustainable wheat industry. This NIFA-funded project includes **two objectives**: (1) to develop a draft SWG genome assembly for genome-specific markers; and (2) to construct an SWG chromosome library in wheat consisting of 14 wheat-SWG addition lines. We have developed a draft assembly of the SWG genome and 127 SWG-specific markers using the assembly and established a GISH procedure to distinguish the two subgenomes. A total of 55 wheat plants carrying one or two SWG chromosomes have been selected by genotyping large backcross populations and a complete set of 14 wheat-SWG chromosome addition lines have been identified using GISH analysis of the plants. In addition, we localized several agriculturally important traits to SWG chromosomes, including the solid stem to chromosome 3J₁, waterlogging tolerance to 1J₁, and WSMV resistance to

2J₁. With these results, we are one step closer to our goal to transfer the biotic stress resistance and abiotic stress tolerance from SWG to wheat.

PE0914: Wheat, Barley, Oat, and related

Physiological and Transcriptomic Characterization of Superior Waterlogging Tolerance in the Sea Wheatgrass-Derived Germplasm of Wheat

Wenqiang Li, Ghana S. Challa and **Wanlong Li**, South Dakota State University, Brookings, SD

Waterlogging is an increasing threat to world and US agriculture as climate change projects more floods. The primary effect of waterlogging comes from the hypoxic stress to root growth and development. Very little is known about the physiological mechanisms underlying waterlogging tolerance in dryland crops like wheat. We recently identified superior waterlogging tolerance in sea wheatgrass (SWG; *Thinopyrum junceiforme*), a distant relative of wheat and a relatively untapped source for wheat improvement. Our long-term goal to understand the genetic mechanisms mediating hypoxia response and to improve waterlogging tolerance in the dryland cereals. This NIFA-funded project includes two objectives: 1) determine the morphological and physiological features of the SWG-derived waterlogging tolerance, and 2) identify hypoxic response genes and pathways by profiling the waterlogging tolerance-dependent root transcriptomes and proteolysis assay. Comparative analysis of the wheat parent and wheat-SWG amphiploid indicated that the formation of aerenchyma and barrier to radial oxygen loss does not contribute significantly to the waterlogging tolerance, but the continuous formation of adventitious roots and prolonged life span of the adventitious root tips underlie the SWG-derived waterlogging tolerance. Analysis of the root tip transcriptome indicated induction of ERF VII TFs and NO production pathways.

PO0915: Wheat, Barley, Oat, and related

Improved Genome Assembly for *Aegilops tauschii* with the Aid of Optical Maps and Whole Genome SMRT Sequencing

Le Wang, Tingting Zhu, Karin R. Deal, Jan Dvorak and Ming-Cheng Luo, Department of Plant Sciences, University of California, Davis, Davis, CA

Aegilops tauschii, the donor of the bread wheat D subgenome, harbors many important resistance genes for abiotic and biotic stresses, and thus is an important genetic resource for wheat breeding. The reference-quality genome sequence for *Ae. tauschii* acc. AL8/78 (Aet v4.0) was a milestone in the assembly of such a large and complex genome. Yet there is still room for improvement by correcting minor assembly errors and closing gaps in Aet v4.0. Nick-based and Direct Label and Stain (DLS) whole genome Bionano optical maps were used to revise scaffolds which had ambiguous orientations or were incorrectly placed and to re-estimate gap sizes. Then, contigs assembled from whole genome PacBio SMRT reads were used for gap-closing guided by optical maps. The improved genome assembly of *Ae. tauschii* acc. AL8/78 (Aet v5.0) was generated. The total effective length (excluding Ns) of Aet v5.0 increased by 9.69 Mb compared to Aet v4.0. The number of gaps (52,910) and total gap length (69.3 Mb) in Aet v5.0 were reduced compared to the corresponding values for Aet v4.0 (91,809 and 82.8 Mb, respectively). This work is a part of the NSF-funded Project IOS-1238231 to generate a reference sequence for the genome of *Ae. tauschii* (<http://aegilops.wheat.ucdavis.edu/ATGSP/>)

PE0916: Wheat, Barley, Oat, and related

Exome Capture in Wheat: What We Can Learn about Diversity in the D-Genome

Marcus A. Hooker, Washington State University, Pullman, WA and Deven R. See, USDA-ARS/Washington State University, Pullman, WA

Crop domestication is known to result in a decrease in genetic diversity as only plants demonstrating desirable characteristics are chosen for breeding, resulting in a founder effect. Additionally, polyploidization acts as another founder effect and can result in a decrease in genetic diversity when the polyploid offspring are no longer able to interbreed with their diploid progenitors. Both cases can be seen in hexaploid bread wheat (*Triticum aestivum*), which went through two independent polyploidization events and resulted in an overall lack of genetic diversity, particularly on the D-genome that resulted from the most recent polyploidization event. Here we attempt to use

exome capture data to explore diversity on the D-genome of wheat. In particular, we mapped known markers developed from the Wheat 90K SNP dataset to our more extensive exome capture dataset to uncover new and informative markers and aid in genotyping projects that inform breeding practices. Additionally, we used the exome data from 18 synthetic wheat samples to understand how much more diversity is available in synthetic wheats over common wheats by creating a phylogenetic tree from the SNPs on the D-genome of all synthetics plus Chinese Spring. This approach allowed us to understand which sources were most distantly related to Chinese Spring and which sources yielded the most informative SNPs. This will inform future attempts to increase D-genome diversity by helping to capture the maximum amount of diversity possible and to incorporate new and desirable traits into bread wheat.

PO0917: Wheat, Barley, Oat, and related

Quantifying Early Root Growth in Wheat Using the LemnaTec Phenocenter Imaging System

Patricia Okubara, USDA-ARS, Pullman, WA and **Brian Bellinger**, USDA ARS, Pullman, WA

Rhizoctonia root rot, caused by the soilborne fungal pathogen *Rhizoctonia solani* AG-8, is a chronic disease of wheat and barley in the Pacific Northwest, USA, Australia and other parts of the world. Genetic resistance to the pathogen has been reported, but the mechanism(s) of action of resistance and underlying resistance genes remain elusive. In wheat, reduced rates of early root growth have been observed in resistant genotypes compared to susceptible genotypes. This observation was obtained by manual measurements of root length of plated seedlings every 4 hours over a 16-hour period, and confirmed by total root length measurements taken at 48 h post-planting using the pixel-counting program WinRHIZO. Using this approach, root lengths of susceptible and resistant genotypes were binned into length categories and plotted against time. This method was laborious and at best semi-quantitative. Recently, we have adapted the LemnaTec Phenocenter (PC-50T Lab Scanalyzer), an automated imaging system, to monitor root growth of wheat seedlings in Petri plates at 4-hour intervals around the clock. The system processes 84 individual seedlings per experiment, or 14 seedlings for each of 6 genotypes. The imaging system also provides an estimated time of root emergence. A description of the automated workflow and analysis pipeline, and progress in quantification of root growth will be reported.

PE0918: Wheat, Barley, Oat, and related

A Major Root Angle QTL in Durum Wheat Improves Yield in Drought and Crown Rot Environments

Samir Alahmad¹, Khaoula EL Hassouni², Eric Dinglasan³, Jack Christopher⁴, Jason Able⁵, Filippo Maria Bassi⁶, Kai Voss-Fels¹ and Lee Hickey⁷, (1)The University of Queensland, Brisbane, QLD, Australia, (2)International Center for Agricultural Research in the Dry Areas, Rabat, Morocco, (3)Centre for Plant Science, Queensland Alliance for Agriculture and Food Innovation, University of Queensland, Australia, Brisbane, QLD, Australia, (4)The University of Queensland - Leslie Research Facility, Toowoomba, QLD, Australia, (5)The University of Adelaide, Adelaide, SA, Australia, (6)ICARDA, Rabat, Morocco, (7)Centre for Plant Science, Queensland Alliance for Agriculture and Food Innovation, University of Queensland, Brisbane, Australia

Durum wheat (*Triticum turgidum* L. ssp. *durum*) production can experience significant yield losses due to crown rot (CR) disease. Losses are usually exacerbated when disease infection coincides with terminal drought. Durum wheat is very susceptible to CR and resistant germplasm is not currently available in elite breeding pools. Deploying physiological traits for drought adaptation such as enhanced root system architecture to reduce water stress, might minimise losses due to CR infection. A subset of lines from a nested association mapping population was evaluated for stay-green traits, CR disease incidence and yield in field experiments, as well as root traits under controlled conditions. Weekly measurements of normalized difference vegetative index (NDVI) in the field enabled modelling of the senescence pattern and calculation of stay-green traits for each genotype. Genome-wide association studies using 2,541 high quality polymorphic DArTseq markers identified a major QTL on 6A (*qSRA-6A*) and 6B (*qCR-6B*) underpinning seminal root growth angle and CR tolerance, respectively. Haplotype analyses identified allelic variants with favourable impact on yield under drought and CR environments. Results of this study highlight the value of combining above- and below-ground physiological traits to enhance yield potential. We anticipate these

insights will assist breeders to design improved durum varieties that mitigate production losses due to water deficit and CR.

PO0919: Wheat, Barley, Oat, and related

Genetic Variation in Nitrogen Partitioning and Grain Protein Content in Wheat

Mamoru Okamoto, Margaret Kirika and Julie Hayes, University of Adelaide, Urrbrae, SA, Australia

Grain protein content (GPC) is a key attribute for grain quality in wheat. GPC is also a primary criterion of wheat grain classification in some countries, such as Australia, the USA, and Canada, influencing growers' revenue and trading values. However, less genetic options are available to produce stabilised GPC in wheat breeding. To date, agronomic approach is the main driver to control GPC by applying nitrogen (N) fertilizer. The research aim is to identify genetic variation in GPC and N partitioning in different tissue types in wheat. Sixteen wheat varieties, selected for contrasting GPC, were grown under two N regimes in a semi-hydroponic system to see N response in GPC and N distribution. When wheat plants were grown in low-N (0.5mM N) condition, 70-80% of total N was accumulated in grains at maturity, whereas high-N (5mM N) grown plants dropped the N proportion in grains down to 50 -70% range. The excess amount of N was stored in stem and non-grain tissues in spikes. Some Mexican varieties maintained higher proportion of N in grains under high-N condition, indicating better efficiency in N utilisation. GPC was in the range of 13 – 23% and 11 – 18% under high and low N condition, respectively. Interestingly, the highest and lowest GPC varieties were same in both N conditions, whereas GPC of the other fourteen varieties varied. Potential genetic approaches to improve GPC in wheat will be discussed.

PE0920: Wheat, Barley, Oat, and related

Genome Wide Association Mapping of Glume Color in A-genome Wheat Species

Laxman Adhikari¹, Shuangye Wu¹, John Raupp¹, Jesse Poland¹, Simon G. Krattinger² and Michael Abrouk², (1)Department of Plant Pathology, Kansas State University, Manhattan, KS, (2)King Abdullah University of Science and Technology, Thuwal, Saudi Arabia

Glume color in A-genome species, the A-genome progenitor of wheat (*Triticum aestivum* L.), can be an important trait to classify the species and differentiate accessions for core collections. Glume color might also be an important trait influencing grain properties. However, the genetic basis of glume coloration (GlmCol) in these species is not illustrated yet. We performed an association analysis of GlmCol in three A-genome species (*T. urartu*, *T. monococcum* subsp. *monococcum*., and *T. m. subsp. aegilopoides*) using genotyping-by-sequencing (GBS) SNP to understand the genetic architecture of the trait in these different species. Nine 96-plexed GBS libraries with *Pst*I-*Msp*I were constructed for 848 A-genome accessions, including 559 *T. m. subsp. aegilopoides*, 172 *T. urartu*, and 117 *T. m. subsp. monococcum*. The raw sequence data (pair-end 150 bp) were processed using the TASSEL5 GBSv2 pipeline, where reads were aligned to an *urartu* reference (tu2.0). The filtered SNPs and binary coded (0 = white, 1= color) phenotype were tested for genome wide association (GWAS) within each species using GAPIT, and the result was verified using rrBLUP mixed model. The top GWAS hit for all three species were observed on the short arm of chromosome (Chr) 1 at about 5 Mb. Further, when a functional query sequence of a wheat *MYB* aligned to the tu2.0 using BLAST, the best hit also observed at ~ 5Mb of Chr 1. These results indicate that the wheat *MYB* ortholog at Chr 1 could be a potential candidate gene for GLmCol in A-genome species.

PO0921: Wheat, Barley, Oat, and related

Identification of a Transcriptional Repressor Underlying B1 Awn Suppression in Wheat

Noah DeWitt¹, Mohammed Guedira², Edwin Lauer¹, Jose Martin Sarinelli¹, Pryanka Tyagi¹, Daolin Fu³, QunQun Hao³, Paul Murphy¹, David Marshall⁴, Alina Akhunova⁵, Eduard Akhunov⁶, Katherine Jordan⁷ and Gina Brown-Guedira^{1,4}, (1)North Carolina State University, Raleigh, NC, (2)Department of Crop & Soil Science, Raleigh, NC, (3)University of Idaho, Moscow, ID, (4)USDA-ARS, Raleigh, NC, (5)Integrated Genomics Facility, Kansas State University, Manhattan, KS, (6)Department of Plant Pathology, Kansas State University, Manhattan, KS, (7)Kansas State University, Department of Plant Pathology, Manhattan, KS

Awns are stiff, hair-like structures which grow from the lemmas of wheat (*Triticum aestivum* L.) and other grasses that contribute to photosynthesis and play a role in seed dispersal. Variation in awn length in domesticated wheat is primarily controlled by three major genes, most commonly the dominant awn suppressor Tipped1 (*B1*). This study identifies a transcription repressor responsible for awn inhibition at the *B1* locus. Association mapping was combined with analysis in bi-parental populations to delimit *B1* to a distal region of 5AL co-localized with QTL for number of spikelets per spike, kernel weight, kernel length, and test weight. Fine-mapping located *B1* to a region containing only two predicted genes, including C2H2 zinc finger transcriptional repressor TraesCS5A02G542800 up-regulated in developing spikes of awnless individuals. Deletions encompassing this candidate gene were present in awned mutants of an awnless wheat. Sequence polymorphisms in the *B1* coding region were not observed in diverse wheat germplasm while a nearby polymorphism was highly predictive of awn suppression.

PE0922: Wheat, Barley, Oat, and related

Profiling of Wheat Stem Fructans from a Canadian Perspective

Letitia Da Ros¹, Daiqing Huang¹, Parul Jain¹, Lanette Ehman¹, Bahati Chisala-Nyekanyeka¹, Jenna Van Parys¹, Ron E. Knox², Richard Cuthbert², Yuefeng Ruan², Jennifer Mitchell Fetch³, Andrew Burt³, Santosh Kumar³ and Allan Feurtado¹, (1)National Research Council Canada, Saskatoon, SK, Canada, (2)Swift Current Research and Development Centre, Agriculture and Agri-Food Canada, Swift Current, SK, Canada, (3)AAFC BRDC, Brandon, MB, Canada

Fructan polymers are the major form of carbohydrate reserve in wheat (*Triticum aestivum*) and accumulate in stems after anthesis. These stem reserves contribute to grain development, especially during periods of environmental stress when photosynthetic carbon assimilation has been compromised. Knowledge of variation in stem fructan concentrations of current Canadian wheat cultivars needs to be established to aid in optimizing grain yield. Stem fructan profiles of fourteen cultivars, at eighteen days post-anthesis and at maturity, were quantified using high-performance anion-exchange chromatography. Total fructan concentrations across cultivars ranged from 9.8% to 23.3% DW, demonstrating the large variation present within current germplasm. 200 double haploid (DH) lines, generated from a cross between hard red spring wheat cultivars with differing fructan profiles, were further analyzed and used in multi-QTL mapping to identify fructan-related QTL. A strong QTL on chromosome 7A was found at a location harbouring a cluster of known fructan biosynthesis genes. The sequence variation and expression patterns of fructan synthesis genes in the QTL interval, such as 1-SST, 1-FFT, 6-SFT and fructan exohydrolases 1-FEH and 6-FEH, were further compared to elucidate why fructan profiles differed between the parental genotypes. Gene specific KASP markers were developed based on the sequence polymorphisms and a panel of Canadian wheat cultivars were genotyped with the markers. The synthesis of these results will help guide future breeding efforts in minimizing stress-related yield reductions in wheat.

PO0923: Wheat, Barley, Oat, and related

The Effects of *RhtB1b* and *RhtD1b* Reduced Height Alleles to the HEAT-Stress Sensitivity of Hexaploid Wheat

Diána Makai, Edit Mihók, András Cseh, Tibor Kiss, Ildikó Karsai, Eva Darkó, Katalin Jager and **Adel Seps**, Centre for Agricultural Research, Agricultural Institute, Martonvásár, Hungary

High-yielding semi-dwarf wheat varieties predominate worldwide wheat production and account for more than 95% of today's global wheat acreage. Widely used in wheat breeding are *RhtB1b* and *RhtD1b* mutant alleles that confer gibberellin insensitive dwarfism. Dwarf growth is manifested in 20% plant height reduction and 5-10% increase in yield. Yield increase and resistance to lodging are due to the improved biomass partitioning and reduced stem elongation. Extreme temperatures imposed by global warming presently affect annual wheat grain yield and in the future, each °C of increase in the global mean temperatures is expected to decline grain production by 6%. By 2050 maximum daily temperatures are expected to exceed a threshold of 30 °C in spring, when wheat plants are in a highly sensitive reproductive stage. As almost all commercial wheat variety carry dwarfing alleles, it is reasonable to ask how the altered dwarfing patterns affect the fertility of semi-dwarf and dwarf wheats. The objective of our study was to investigate how heat stress applied at an early developmental stage affects the fertility of a wheat near isogenic line set carrying different dwarfing alleles ('Maris Huntsman' wild type, *Rht-B1b*, *RhtD1b* and double mutant, JIC, Norwich, UK). Experiments were carried out in controlled environment cabinets in the Martonvasar

Fitotron Facility (Hungary) and were transferred to stress cabinets at an early developmental stage, when the mean spike reached late meiotic interphase (approx. Zadok's 41). Fertility and yield was manually quantified for the apical, central and basal part of each spike. We show that a short-term elevated temperature (30°C, 24h) significantly reduces the fertility of the double mutant line. Loss of fertility was manifested at the main spike and the whole plant level. Within the main spikes middle regions were the most affected by the elevated temperature. We analysed the expression levels of *Rht-B1b*, *RhtD1b* mutant- and wild type alleles and expressions of genes involved in GA biosynthesis within the leaves and anthers of the near isogenic lines subject to control and elevated temperatures. Our result showed that heat stress significantly affected expressions of both wild and mutant *Rht* alleles and also the final step of GA biosynthesis presumably leading to increased bioactive GA levels.

The research leading to the present results has received funding from the Hungarian Scientific Research Fund (NKFIH, proposal ID FK 124 266) and the Hungarian Academy of Sciences (János Bolyai Research Scholarship).

PE0924: Wheat, Barley, Oat, and related

Solving the *Rht18* Enigma

Pavla Navrátilová¹, Jan Vrána², Brett Ford³, Jaroslav Doležel¹, Wolfgang Spielmeier⁴ and **Hana Šimková²**, (1)Institute of Experimental Botany, Olomouc, Czech Republic, (2)Institute of Experimental Botany, Centre of the Region Haná for Biotechnological and Agricultural Research, Olomouc, Czech Republic, (3)CSIRO, Canberra, Australia, (4)CSIRO Agriculture & Food, Canberra, Australia

The regulation of plant height formed the basis of the Green Revolution, which was associated with major yield increases in wheat and rice due to improved lodging resistance and better allocation of assimilates to grain growth. Proteins encoded by semidwarfing genes regulate the plant height by lowering the bioactive content of gibberellins (GAs) or by inhibiting GA signalling. Previously generated semidwarf durum wheat cultivar Icaro, bearing a height-affecting locus *Rht18* in chromosome 6A, has been characterized as a GA-sensitive mutant with increased degradation of GA precursor, caused by significant overexpression of *GA2-oxidase A9*. As nucleotide sequences of the gene and its adjacent regions were identical in the mutant Icaro and the original tall durum wheat variety Anhinga, we hypothesize that the increased gene expression could result from disrupted transcriptional regulation involving non-coding DNA and/or epigenetic modification.

Aiming to reveal the causative element of the overexpression, we approached the wider *GA2oxA9* region in Icaro and Anhinga by nanopore and bisulfite sequencing of flow-sorted 6A chromosomes, analysis of chromatin accessibility, local profiling of histone modifications and Chromatin Conformation Capture-based analysis (4C-seq) to identify *GA2oxA9*-related chromatin interactions. Combination of these techniques provides comprehensive information at both genetic and epigenetic levels with prospects to reveal long-distance regulatory elements inducing the semidwarf Icaro phenotype.

This work has been supported by ERDF project No. CZ.02.1.01/0.0/0.0/16_019/0000827.

PO0925: Wheat, Barley, Oat, and related

Mapping and Characterization of Yield Component Traits in Cultivated Emmer and Durum Wheat

Amanda R. Peters¹, Andrew J. Green¹, Steven S. Xu² and Justin D. Faris², (1)North Dakota State University, Fargo, ND, (2)USDA-ARS, Fargo, ND

To date, numerous studies have identified yield component genes in hexaploid (*Triticum aestivum*) wheat; however, fewer yield evaluation studies have been performed in tetraploid wheat. A potential source for allelic variation for wheat improvement is cultivated emmer (*T. turgidum* ssp. *dicoccum*). Here we evaluated a recombinant inbred line population (BP025) derived from crossing the durum (*T. turgidum* ssp. *durum*) wheat variety Ben by the cultivated emmer accession PI 41025. The BP025 population was grown under field conditions at Prosper, North Dakota in the summers of 2017, 2018, and 2019. During the 2017 and 2018 seasons, 50 unique QTL have been identified for yield component traits such as spikelets per spike, kernels per spike (KPS), grain weight per spike (GWS), thousand kernel weight, and seed morphology traits. A significant QTL on chromosome 2B for KPS and GWS contributed up

to 14% and 11% of the phenotypic variation for KPS and GWS, respectively. Currently, this QTL spans a 38 Mb region at the 445-483 Mb position on the durum Svevo reference sequence. We developed 24 markers for the target region and mapped them in F₃ homozygous recombinant individuals. The F₄ plants from each of the recombinant families will be phenotyped to narrow the target region through high-resolution mapping. The cloning of this QTL will provide fundamental knowledge regarding the genetic mechanisms controlling yield and provide additional tools for breeders to expedite the development of high-yielding varieties.

PE0926: Wheat, Barley, Oat, and related

Deletion of the Glu-B3 Low Molecular Weight Glutenin Subunit Locus in an Elite Bread Wheat Variety

Debbie Laudencia-Chingcuanco, Gracie Benson-Martin and Jeanie Lin, USDA-ARS, Western Regional Research Center, Albany, CA

The unique visco-elastic property of dough that is important in processing wheat-derived food is determined mainly by the gluten proteins. Gluten is composed primarily of two groups of seed storage proteins known as glutenins and gliadins. Each member of these protein families vary in their effect on dough quality. To understand the role of the individual protein components of the gluten in dough quality we screened a fast-neutron radiation-mutagenized population of Summit, a commercial bread wheat variety, for lines with altered glutenin and gliadin protein profiles. Here, we report the identification and initial characterization of twelve lines with deletions in the *Glu-B3* locus. The mutant lines exhibit a deficiency in a major low molecular weight glutenin subunit (LMW-GS) protein that is similar to the Aroona *Glu-B3-3* allele, an s-type LMW-GS protein. The induced mutations in these lines as revealed by assays using the iSelect Illumina Wheat 90K SNP arrays, ranged from less than 0.01 cM interstitial deletion to the loss of the entire 1B chromosome. Dough derived from *Glu-B3* deficient lines exhibited slightly weaker dough strength and elasticity relative to wild type. The absence of a major LMW-GS protein increased the time to reach maximum resistance during dough development. We show evidence that the set of *Glu-B3* deletion mutants also has deletions in the *Gli-B1* locus that encode omega and gamma-gliadins, proteins that play a role in inducing gluten-related disorders in humans.

PO0927: Wheat, Barley, Oat, and related

Sequence Based Genotyping for Glutenin Genes of Wheat

Emily Delorean¹, Cameron Amos¹, Carlos Guzman², Itria Maria Ibba³, Curtis J Pozniak⁴ and Jesse Poland¹, (1)Department of Plant Pathology, Kansas State University, Manhattan, KS, (2)Universidad de Córdoba, Spain, (3)International Maize and Wheat Improvement Center (CIMMYT), Mexico, (4)University of Saskatchewan, Saskatoon, SK, Canada

The quality of products produced from wheat flour depends largely on gluten forming proteins. Gluten forming proteins are comprised of the high-molecular weight (HMW) glutenins, the low molecular weight (LMW) glutenins, and the gliadins. HMW and LMW glutenin proteins form inter- and intra-molecular bonds to create a large matrix that contributes to the viscoelastic properties typical of wheat dough. Depending on the glutenin alleles constituting the gluten matrix, the technological properties of the wheat dough change, influencing the end-use application of the flour. Therefore, breeding programs must continually monitor the end-use quality of germplasm to ensure consumer acceptance. Given that HMW glutenins are major determinants of the end-use quality and the most widely characterized, wheat breeding programs often use protein based gels or PCR markers to select for the preferred alleles. In this research, we used the extensive sequence information from the 10+ Wheat Genome Project coupled with whole genome sequencing of 95 CIMMYT wheat varieties to identify nucleotide differences between HMW glutenin alleles. These differences were captured as k-mers and diagnostic k-mers identified. We then designed a user-friendly bioinformatics pipeline that searches sequencing data of breeding lines for these diagnostic k-mers and determine the allelic state at the HMW glutenin loci. This approach has the potential to offer a low cost, high-throughput sequence-based alternative to gel methods for gluten genotyping in breeding programs.

PE0928: Wheat, Barley, Oat, and related

Identification of a Candidate Gene for a QTL for Spikelet Number Per Spike on Wheat Chromosome Arm 7AL by High-Resolution Genetic Mapping

Saarah Noriko Kuzay, University of California, Davis, CA

A better understanding of the genes controlling differences in wheat grain yield components can accelerate the improvements required to satisfy future food demands. In this study, we identified a promising candidate gene underlying a quantitative trait locus (QTL) on wheat chromosome arm 7AL regulating spikelet number per spike (SNS). We used large heterogeneous inbred families (>10,000 plants) from two crosses to map the 7AL QTL to an 87-kb region (674,019,191-674,106,327 bp, RefSeq v1.0) containing two complete and two partial genes. In this region, we found three major haplotypes that were designated as H1, H2 and H3. The H2 haplotype contributed the high SNS allele in both H1 x H2 and H2 x H3 segregating populations. The ancestral H3 haplotype is frequent in wild emmer (48%) but rare (~1%) in cultivated wheats. By contrast, the H1 and H2 haplotypes became predominant in modern cultivated durum and common wheat, respectively. Among the four candidate genes, only TraesCS7A02G481600 showed a non-synonymous polymorphism that differentiated H2 from the other two haplotypes. This gene, designated here as WHEAT ORTHOLOG OF APO1 (WAPO1), is an ortholog of the rice gene ABERRANT PANICLE ORGANIZATION 1 (APO1), which affects spikelet number. Taken together, the high-resolution genetic map, the association between polymorphisms in the different mapping populations with differences in SNS, and the known role of orthologous genes in other grass species suggest that WAPO-A1 is the most likely candidate gene for the 7AL SNS QTL among the four genes identified in the candidate gene region.

PO0929: Wheat, Barley, Oat, and related

Counting on Crossovers: Fine-Mapping a Kernel Weight and Morphology Gene in Wheat

Ella R Taagen and Mark E. Sorrells, Cornell University, Ithaca, NY

The kernel weight and shape of wheat (*Triticum aestivum* L.) are valued traits due to their association with yield and milling quality. Recent advancement of the wheat reference genome assembly and genome editing tools can help facilitate the characterization of genes underlying quantitative trait loci (QTL) for yield components. A QTL contributing to kernel weight and width was identified and fine-mapped on chromosome 5A in the Synthetic W7984 × Opata M85 (SynOp) doubled haploid (DH) and recombinant inbred line (RIL) mapping populations. F_{6,6} heterogeneous inbred families (HIFs) were developed from SynOpRIL lines segregating within the QTL flanking markers which improved the resolution of the causal gene variant to two neighboring regions on chromosome 5A. A days-post-anthesis (DPA) time course experiment assessed the developing kernel weight and morphology and concluded that after 10 DPA HIF kernels with the Opata allele at the 5A QTL are significantly heavier and wider than HIF kernels with the W7984 allele. Relying on large populations over many generations to detect crossovers and capture finer resolution of the QTL is resource limiting. Our attention has now turned to genome editing tools, including controlled recombination, to facilitate characterization of the underlying causal grain morphology gene. Our ultimate goal is to clone this beneficial grain size and shape gene and deploy it in commercial varieties preferred by domestic and international wheat-growers and mills.

PE0930: Wheat, Barley, Oat, and related

Developing an Immunoassay for Late Maturity α -Amylase (LMA) and Preharvest Sprouting (PHS).

Amber Hauvermale¹, Andy McCubbin², Michael Pumphrey¹ and Camille M. Steber³, (1)Department of Crop and Soil Sciences, Pullman, WA, (2)School of Biological Sciences, Pullman, WA, (3)USDA-ARS Wheat Genetics, Quality, Physiology and Disease Research Unit, Pullman, WA

In wheat, the aleurone layer is a single layer of specialized cells surrounding the starchy endosperm. The aleurone layer produces the enzyme α -amylase in response to the plant hormone gibberellin (GA) during germination. Upon induction, α -amylase is secreted into the starchy endosperm where it cleaves starch molecules to fuel germination. While α -amylase is essential for germination, high expression outside of a normal germination program may negatively impact flour quality, cause poor baking end-use quality, and result in significant financial losses to wheat producers globally. Two genetic factors that lead to high α -amylase expression outside of a normal germination program are late maturity α -amylase (LMA), occurring during grain filling in response to a cold shock, and

preharvest sprouting (PHS) occurring after seed maturation in response to a rain event. The Hagberg-perten Falling numbers test (FN) measures starch damage caused by α -amylase and other enzymes in the flour. Sound flour has a FN above 300 while compromised flour has a FN below 300. Although both LMA and PHS may contribute to a low FN, they do so through different molecular mechanisms, and the effect on end-use quality may not be equivalent. However, the Falling Numbers test cannot directly determine differences in LMA and PHS physiology or the impacts to end-use quality. To address the limitations in the existing technology an immunoassay platform using monoclonal antibodies to wheat α -amylases and other germination specific enzymes is being developed as a more reliable test for detecting LMA and PHS.

PO0931: Wheat, Barley, Oat, and related

Molecular Mapping of High Amylose Starch in Wheat Using GBS Approach

Ankita Mishra^{1,2}, Vinita Sharma^{1,3}, Mohammed Saba Rahim¹, Humira Sonah¹, Shrikant Subhash Mantri¹, Tilak Raj Sharma¹ and Joy K Roy¹, (1)National Agri-Food Biotechnology Institute (NABI), Mohali, India, (2)Panjab University, Chandigarh, India, (3)Indian Institute of Science Education & Research (IISER), Mohali, India

The present lifestyle has raised the prevalence of chronic diseases including obesity and diabetes as major health concern globally. Wheat plays a central role in diet globally, provides 20% of caloric consumption (particularly from starchy endosperm). Starch is a major constituent of whole grain comprising of highly linear glucose polymer amylose (25%) and mostly branched polymer amylopectin (75%). High amylose starch has the potential to improve human health and lower the risk of chronic diseases. This study aimed to understand the genetic basis of high amylose content in wheat (*Triticum aestivum* L.). A bi-parental (F_{2:3}) mapping population derived from 'WH 1105' (high yielding Indian wheat variety) and 'TAC 75' (high amylose EMS mutant line) was used for genetic mapping. Genotyping by sequencing (GBS) markers facilitated the identification of 14 major QTLs associated with trait, individually explaining phenotypic variance from 10-30%. The QTL on 7A was consistent in F₂ and F₃ generation, spanning a region of 31Mb harbouring 236 high confidence protein coding genes. GBSSI, a key enzyme for amylose biosynthesis, was identified within a major QTL (7A). A total of 12 non-synonymous mutations identified within the *GBSSI* gene including a novel transition mutation, which led to amino acid substitution. The enhanced activity and expression profile of the enzyme indicated its contribution towards the high amylose phenotype. Three high amylose mutant lines identified in the present study have low glycemic index (30% lower in glucose response than parent variety). Efforts are being made to commercialize the low glycemic wheat mutant lines.

PE0932: Wheat, Barley, Oat, and related

SNPs Based Genetic Diversity Analysis of CIMMYT Originated Wheat Germplasm to Identify High Resistant Starch Genotypes for Type-2 Diabetic Patients

Zaheer Ahmed, University of Agriculture Faisalabad, Faisalabad, Pakistan

Starch provides more than 70% daily caloric intake worldwide. This starch is readily digested to glucose leading to rapid increase in glycaemic index (GI), rendering an individual susceptible to type-2 diabetes. In contrast, resistant starch (RS), exhibits naturally or artificially modified structure. Increased RS uptake in daily food can significantly help in preventing colon cancer, early occurrence of type-2 diabetes and improves cardiovascular health. Synthesis of RS is genetically controlled. Therefore, present study was designed to determine the genetic diversity in CIMMYT originated, sixty wheat lines and to understand the mechanism of RS synthesis to identify suitable variants for the development of RS enriched wheat lines. SNPs targeting starch synthesizing enzymes were used to determine the genetic diversity among genotypes. Results indicated that genotypes exhibited variations for suitability to breeding program aimed at increasing RS content. However, this genetic diversity was narrow due to more emphasis on selection for yield. Increased, RS content was associated with altered physico-chemical starch properties. It was found that proportion of B-granules and altered granule morphology was highly correlated with increased RS. Our results also indicated that granule bound starch synthase (GBSS) and starch branching enzymes (SBEs) play major role in increasing RS content but genetic diversity targeting SBEs was found more effective. Alteration in protein complexes compared to wild type also added in increased RS content. This study first of its kind, shows that CIMMYT-originated Pakistani wheat germplasm exhibits genetic diversity which can be significantly used to develop wheat material with increased RS content.

PO0933: Wheat, Barley, Oat, and related

Analysis of Organellar Genome Variants and Their Role in Evolution, Development and Defense Response in Wheat

Keshav Sharma¹, Katie Liberatore², Marisa Miller³ and Shahryar Kianian¹, (1)USDA/ARS, Saint Paul, MN, (2)Calyxt, Roseville, MN, (3)Pairwise, Durham, NC

Organelles play important roles in physical and physiological functions of all plants. Our study is aimed at the analysis of organellar genome diversity and their variation in wheat and related wild species. Due to polyploidization events leading to the formation and domestication of tetraploid pasta and hexaploid bread wheat, the organelle genome variation was severely reduced in modern wheat cultivars. This lack of variability could limit overall flexibility and adaptability of plants to drastic shifts in stress factors. To better understand the organelle genome diversity, their role in development and evolution, various wild, euplasmic and alloplasmic lines were studied at the sequence level. Here, we report on the diversity of the organellar genome in *Triticum spp.* and *Aegilops spp.* and their possible impact on plant development. Organellar DNA of fourteen *Aegilops spp.* and several *Triticum spp.* such as *Triticum durum*, *Triticum dicoccoides* and *Triticum aestivum* were subjected to PacBio and Illumina sequencing. Resulting sequences were separated into mitochondrial and chloroplast sequences, trimmed and *de novo* assembled. They were then analyzed for within species as well as between species variability to identify the impact of observed changes. This information is critical for better analysis of nuclear-cytoplasmic crosstalk and its impact on plant development and stress response.

PE0934: Wheat, Barley, Oat, and related

Mapping Wheat Quantitative Trait Loci for Fusarium Head Blight Resistance and Agronomic Traits in a Mutant from Jagger

Ruolin Bian, Kansas State University / Agronomy, Manhattan, KS

Wheat Fusarium head blight (FHB) causes significant yield losses and is one of the most destructive diseases of wheat. To explore new sources of resistance, we identified one resistant line (1095EMSMut) by phenotyping EMS-treated FHB-susceptible Jagger lines. FHB tests in greenhouse showed that 1095EMSMut showed significantly lower disease severity than Jagger. Also, 1095EMSMut has phenotypic mutations in several agronomic traits including plant height, spikelet number and tillering number etc. We hypothesize that the phenotypic change from FHB susceptibility to resistance is due to a mutation in a novel FHB resistance gene that changed a susceptible allele in Jagger to a resistance allele in the mutant. To identify the novel FHB resistance gene and quantitative trait loci (QTLs) for these mutated agronomic traits, we developed 156 recombinant inbred lines (RILs) of 1095EMSMut x Jagger. The RIL population was phenotyped for FHB resistance and agronomic traits in two greenhouse experiments and genotyped using genotyping-by-sequencing (GBS) markers. A total of 1411 GBS-SNPs were mapped on 21 wheat chromosomes. Using the GBS-SNP map and the phenotyping data of RIL population, we detected one major QTL for FHB resistance on chromosome 4B, two stable QTLs for plant height on chromosome 4B and 2B, two QTLs for spikelet number on 5D and 2B, and two QTLs for tillering number on 4B and 2B. These results lay solid foundation for further fine mapping and map-based cloning of these genes for FHB resistance and other important agronomic traits.

PO0935: Wheat, Barley, Oat, and related

Mining Natural Diversity and NLR-IDs to Engineer Disease Resistance

Janina Tamborski, University of California Berkeley, Berkeley, CA, Daniil Prigozhin, Lawrence Berkeley National Laboratory, Berkeley, CA, Elisha Thynne, The Sainsbury Laboratory, Norwich, United Kingdom and Ksenia Krasileva, Department of Plant and Microbial Biology UC Berkeley, Berkeley, CA

Plants deploy a multitier immune system comprising of plasma membrane localized immune receptors that surveil the extracellular space and soluble cytoplasmic immune receptors that monitor the cell's interior. The intracellular receptors are termed Nucleotide-Binding Leucine-Rich Repeat Receptors (NLRs) and recognize the presence or action of pathogen molecules (effectors) that are secreted into the plant cell to promote pathogen virulence. Recognition of an effector through NLR activation leads to a localized cell death response that is generally referred to as hypersensitive response. Using synthetic biology, we introduce rational modifications in two classes of NLRs: those that directly bind effectors and those NLRs with integrated domains (IDs) which have incorporated plant

protein domains into their otherwise conserved protein structure. Our analysis of allelic variation of direct binding NLRs together with structural modelling have revealed highly variable amino acid residues that we hypothesize are most likely directly involved in effector binding and are therefore under high selective pressure. The modification of these residues should result in altered binding specificity. Conversely, NLRs with integrated domains present the opportunity of exploiting the action of effectors to the disadvantage of the pathogen. We screen native and engineered NLR-IDs for the recognition of effectors that target the corresponding plant protein of the integrated domain.

PE0936: Wheat, Barley, Oat, and related

Rapid Identification of a Novel Wheat Leaf Rust Resistance Gene By Sequencing Contrasting Genotypes in an F₂ Population

Xiangyang Xu¹, Genqiao Li², Amy N. Bernardo³, Guihua Bai⁴, Paul AtAmand⁴ and Brett F. Carver⁵, (1)USDA-ARS, Wheat, Peanut and Other Field Crops Research, Stillwaters, OK, (2)USDA-ARS, Wheat, Peanut and Other Field Crops Research, Stillwater, OK, (3)USDA-ARS-Small Grain Genotyping Lab, Manhattan, KS, (4)USDA-ARS, Manhattan, KS, (5)Oklahoma State University, Stillwater, OK

Leaf rust, caused by *Puccinia triticina* Erikss. (*Pt*), is the most common disease of wheat (*Triticum aestivum* L.) worldwide. Most leaf rust resistance genes have lost effectiveness in the Great Plains of the USA because of the abundant variation in *Pt* populations. Identification of novel leaf rust resistance genes is imperative for sustainable wheat production. To characterize the leaf rust resistance gene in Iranian landrace PI 622111, an F₂ population and 175 F_{2:3} lines derived from PI 622111 × Yuyuan 3 were evaluated for seedling responses to *Pt* race *Pt52-2* (MMPSD). The χ^2 tests indicated a single gene segregated in the F₂ and F₃ populations ($\chi^2_{3:1} = 1.15$, $df = 1$, $p = 0.28$; $\chi^2_{1:2:1} = 2.39$, $df = 2$, $p = 0.30$). Based on F_{2:3} phenotypic data, 12 F₂ plants showing homozygous resistance and 12 F₂ plants exhibiting homozygous susceptibility were sequenced to develop GBS (genotyping by sequencing) markers, leading to the identification of S5B_9219863, positioned at 9,219,863 bp on the short arm of chromosome 5B in the Chinese Spring IWGSC RefSeq v.1.0 reference sequence. All resistant and susceptible plants carried 'A' and 'G' alleles at the S5B_9219863 locus, respectively, indicating that the leaf rust resistance gene in PI 622111 is close to S5B_9219863. A total of 1,445 SSR loci were identified in the terminal region (0-13.5 Mb) of chromosome 5BS, and 48 of them were chosen to develop SSR markers. Linkage analysis delimited the leaf rust resistance gene in PI 622111, designated *Lr622111*, in an interval of 1.1 Mb flanked by *Xstars669* (6.5 Mb) and *Xstars678* (7.6 Mb). *Lr622111* was 0.5 cM proximal to *Xstars669* and 6.1 cM distal to *Xstars678*. *Lr622111* is a new gene differing with *Lr52*, a leaf rust resistance gene located on 5BS, in responses to *Pt* races. *Lr622111* can be used to enhance leaf rust resistance in the Great Plains of the USA.

PO0937: Wheat, Barley, Oat, and related

Single and Multi-Locus Genome-Wide Association Study for Leaf Rust Resistance in Cultivated, Progenitor and Wild Wheat

Fizza Fatima^{1,2}, Brent McCallum³, Colin Hiebert³, Curt McCartney³, George Fedak², Curtis Pozniak⁴, Tara Edwards², Sridhar Ravichandran², Frank You² and Sylvie Cloutier^{1,2}, (1)University of Ottawa, Ottawa, ON, Canada, (2)Ottawa Research and Development Centre, Agriculture and Agri-Food Canada, Ottawa, ON, Canada, (3)Morden Research and Development Centre, Agriculture and Agri-Food Canada, Morden, MB, Canada, (4)Crop Development Centre, University of Saskatchewan, Saskatoon, SK, Canada

Polyploidization, domestication and breeding bottlenecks have reduced genetic diversity in cultivated wheat in comparison to its progenitors and wild relatives, which are now widely recognized as important resources to acquire useful traits. Here, we assessed a collection of 385 accessions, comprising 28 different species of cultivated wheat, synthetic hexaploid wheat, progenitor species, and wild relatives using the wheat 90k genotyping array. Genotype calling and SNP filtering identified 20,448 SNPs, of which, 8,031, deemed high-confidence through their validation using exome capture, mapped onto the *Chinese Spring* reference sequence. Principal component and phylogenetic analyses illustrated the relationships among species and ploidy levels.

A genome-wide association study (GWAS) was conducted for leaf rust severity and reaction ratings against six leaf rust races. Phenotyping for severity was performed in separate field trials for the spring and winter panels for 3-4 years at 2-3 locations in Canada, while the race-specific reactions were conducted in a greenhouse. GWAS was performed using eight single and multi-locus models, identifying 118 unique quantitative trait nucleotides (QTNs) for race-specific infection type, and 102 and 57 QTNs for severity in the spring and winter panels, respectively. Thirty-nine QTNs for race-specific infection types and 37 from severity were located within known disease resistance related genes. Of the total QTNs identified, 21 were within 5Mb of one or more of 65 of the known *Lr* genes mapped. Taken together, this research highlights the potential of the 90k array for inferring population structure and identifying the location of new genes from a diverse germplasm.

PE0938: Wheat, Barley, Oat, and related

Genetic Mapping and Marker Development for Wheat Leaf Rust Resistance Gene *Lr32*

Jyoti Saini Sharma, Curt McCartney, Brent McCallum and **Colin Hiebert**, Morden Research and Development Centre, Agriculture and Agri-Food Canada, Morden, MB, Canada

Leaf rust, caused by the fungal pathogen *Puccinia triticina* Eriks, is the most broadly distributed disease of wheat. Host genetic resistance is a desirable method of disease control. However, leaf rust resistance (*Lr*) genes can lose effectiveness as the pathogen evolves virulence. This erosion of resistance is exacerbated by deploying *Lr* genes singly in cultivars. One strategy to develop cultivars with effective resistance is to identify and incorporate known *Lr* genes that are broadly effective and have had limited use in agriculture, thereby limiting the opportunity to select new virulence against these genes. By using these genes in gene stacks or combinations the duration of their effectiveness is maximized. One such gene is *Lr32* which was identified in *Aegilops tauschii*, is broadly effective, and has not been deployed in cultivars occupying significant acreage. DNA markers are needed to facilitate selection of gene combinations. The goals of this study were to develop a high resolution linkage map and develop SNP markers for *Lr32*. Initial mapping was performed using a doubled haploid (DH) population (n = 244) from the cross Thatcher x BW196R (*Lr32* carrier) to identify co-dominant flanking markers. The flanking markers were used to identify recombinants in a large F₂ population also from the cross Thatcher x BW196R that consisted of ~2000 progeny. There were 106 recombinants identified and analysed with leaf rust and DNA markers. Recombinants were tested with SNP markers derived from a variety of sources that were converted to the KASP assay. *Lr32* was mapped to an interval with a genetic size of 2.7 cM and a physical size of ~30 Mb. In total, 18 KASP markers co-segregated with *Lr32*. These markers are well suited for marker-assisted selection of *Lr32* in wheat breeding programs.

PO0939: Wheat, Barley, Oat, and related

A Wheat Multi-Transgene Cassette Provides Stem and Leaf Rust Resistance in the Field

Oadi Matny¹, Mehran Patpour², Liqiong Xie³, Ming Luo⁴, Soma Chakraborty⁵, Aihua Wang⁵, James Kolmer⁶, Terese Richardson⁷, Dhara Bhatt⁷, Mohammad Hoque⁵, Chris Sorenson⁸, Burkhard Steuernagel⁹, Brande B.H. Wulff¹⁰, Narayana Upadhyaya⁷, Rohit Mago⁷, Sambasivam Periyannan⁷, Roger Freedman¹¹, Evans Lagudah¹², T Lynne Reuber¹³, Brian J. Steffenson¹⁴ and Michael Ayliffe⁷, (1)Department of Plant Pathology, University of Minnesota Twin Cities, St. Paul, MN, (2)Dept. of Agroecology, Aarhus University, Slagelse, Denmark, (3)School of Life Science and Technology, Xinjiang University, Xinjiang, ACT, China, (4)CSIRO, Canberra, Australia, (5)CSIRO Agriculture and Food, Canberra, ACT, Australia, (6)USDA-ARS, St Paul, MN, (7)Commonwealth Scientific and Industrial Research Organisation, Canberra, Australia, (8)Dept. of Agroecology, Aarhus University, Forsøgsvej, Slagelse, ACT, Denmark, (9)The Sainsbury Laboratory and John Innes Centre, Norwich, United Kingdom, (10)John Innes Centre, Norwich, United Kingdom, (11)2Blades Foundation, Evanston, IL, (12)Commonwealth Scientific and Industrial Research Organisation, Canberra, ACT, Australia, (13)Two Blades Foundation, San Mateo, CA, (14)Department of Plant Pathology, University of Minnesota, Saint Paul, MN

Stem rust (*Puccinia graminis* f. sp. *tritici*, *Pgt*) and leaf rust (*Puccinia triticina*, *Pt*) are two of the most important diseases of wheat worldwide. Genetic resistance offers the best means of control for these diseases, but often lacks

durability due to changes in pathogen virulence. To enhance durability of rust resistance in wheat, a multi-transgene gene cassette consisting of four all-stage resistance genes (*Sr22*, *Sr35*, *Sr45* and *Sr50*) and one multi-pathogen adult plant resistance (APR) gene (*Sr55/Lr67*) was introduced in cultivar Fielder using *Agrobacterium*-mediated transformation. Five transgenic lines (three events) with the cassette plus Fielder were grown and inoculated with *Pgt* race QTHJC (virulent on Fielder) in Minnesota in 2018 and 2019. Disease pressure was high in both years as Fielder exhibited disease severities of 60 and 96.%, respectively. Each of the five multi-transgene lines was highly resistant showing only hypersensitive flecks and no pathogen sporulation whatsoever. To assess the level of leaf rust resistance in the transgenic lines in the field, an epidemic was initiated with *Pt* races MNPSD, MBDS and MPPSD in 2019. Area Under Disease Progress Curve (AUDPC) values for the transgenic lines were significantly lower (43 to 117) than for Fielder (456). The reduced leaf rust progression is consistent with *Sr55/Lr67* being functional and providing characteristic partial APR to *Pt*. The pyramiding of all-stage and APR genes in wheat by transformation provides a means to extend the durability of rust resistance and also simplify the breeding process since the cassette segregates as a single locus.

PE0940: Wheat, Barley, Oat, and related

A Stem Rust Resistance QTL Located on Chromosome 5D in Two Contemporary Canadian Spring Wheat Varieties

Firdissa Bokore¹, Ron E. Knox¹, Richard Cuthbert¹, Tom Fetch², Samia Berraies¹, Heather Campbell¹, Amidou N'Diaye³, Curtis J Pozniak³, Andrew G. Sharpe⁴, Colin Hiebert⁵, Curt McCartney⁵, Sachin Kumar⁶, Yuefeng Ruan¹ and Brad Meyer¹, (1)Swift Current Research and Development Centre, Agriculture and Agri-Food Canada, Swift Current, SK, Canada, (2)Brandon Research and Development Centre, Agriculture and Agri-Food Canada, Brandon, MB, Canada, (3)University of Saskatchewan, Saskatoon, SK, Canada, (4)Global Institute for Food Security (U of S), Saskatoon, SK, Canada, (5)Morden Research and Development Centre, Agriculture and Agri-Food Canada, Morden, MB, Canada, (6)Chaudhary Charan Singh University, Meerut, India

Wheat (*Triticum aestivum* L.) stem rust caused by *Puccinia graminis* f. sp. *tritici* (*Pgt*) is a constant threat to crop production. Identifying and mapping stem rust resistance genes allows pre-emptive stacking of the genes through marker assisted breeding. The objective of this study was to map stem rust resistance QTL in three Canadian spring wheat doubled haploid populations - Carberry/AC Cadillac, Stettler/Red Fife, and Carberry/Thatcher. These were evaluated near Swift Current, SK or Brandon, MB Canada for stem rust severity and infection response over multiple years. The populations were genotyped with the Illumina 90K iSelect SNP assay and QTL identified with MapQTL 6. A stable QTL on chromosome 5DL was identified across the three mapping populations. Carberry and Stettler were the sources of resistance. The QTL was significant in nine out of ten population-environments for either disease severity or infection response or both. The phenotypic variation explained by the QTL ranged from 2.1 to 23.2%. The SNP markers *Excalibur_c34793_1260*, *Kukri_c15380_902* and *Kukri_c52028_104* anchored the 5DL QTL across the populations suggesting the QTL is shared between Carberry and Stettler. Markers associated with the QTL are assigned to the long arm of chromosome 5D. The relationship of the QTL to the two stem rust resistance genes, *Sr30* and *Sr53*, located on 5DL is unknown.

PO0941: Wheat, Barley, Oat, and related

Development and Validation of Semi-Thermal Asymmetric Reverse PCR Markers for Ug99-Effective Stem Rust Resistance Genes in Wheat

Jinglun Zhang¹, Yuming Long¹, Baljeet K. Gill¹, Jyoti Saini¹, Qijun Zhang¹, Daryl L. Klindworth², Xiwen Cai³, Timothy L. Friesen², Justin D. Faris² and Steven S. Xu², (1)North Dakota State University, Fargo, ND, (2)USDA-ARS, Fargo, ND, (3)Department of Plant Science, North Dakota State University, Fargo, ND

The Ug99 race group of the stem rust pathogen (*Puccinia graminis* f. sp. *tritici*) is currently considered a serious threat to global wheat production. Identification and deployment of Ug99-effective stem rust resistance (*Sr*) genes have become major areas of focus in many breeding programs around the world. So far, over 20 Ug99-effective genes have been identified in wheat and its relatives. We recently developed a robust and user-friendly PCR-based method named semi-thermal asymmetric reverse PCR (STARP) for genotyping individual SNPs and indels and

initiated a project to develop STARP markers for Ug99-effective genes with high breeding value. We previously developed STARP markers for *Sr13*, *Sr47*, *Sr8155B1*, and two potentially novel genes from an emmer wheat accession. Here, we report the development and validation of diagnostic STARP markers for three *Sr13* alleles, *Sr39*, *Sr43*, and a new *Sr* gene recently introgressed from *Aegilops markgrafii*. Among these new markers, two (*Xrwgshnp38* and *Xrwgshnp39*) were developed based on the functional SNPs of the cloned *Sr13* haplotypes R1 (*Sr13a*), R2 (*Sr13b*), and R3 (*Sr13c*). *Xrwgshnp38* differentiates between R1/R3 (*Sr13a/Sr13c*) and R2 (*Sr13b*), and *Xrwgshnp39* differentiates between R1 (*Sr13a*) and R3 (*Sr13c*). The two markers (*Xrwgshnp38* and *Xrwgshnp39*), along with the previously developed *Sr13* locus-specific STARP marker *Xrwgshnp37*, have been used extensively for detection of *Sr13* alleles in durum and related tetraploid wheat germplasm. The markers for *Sr39*, *Sr43*, and the new *Sr* gene from *Ae. markgrafii* have been used in marker-assisted deployment of these genes into adapted durum and/or bread wheat germplasm.

PE0942: Wheat, Barley, Oat, and related

Exome Sequencing-Based Fine Mapping of Stem Rust Resistance Genes in Soft Red Winter Wheat Cultivar AGS2000 (*Triticum aestivum* L.)

Luis Rivera-Burgos¹, Mohammed Guedira¹, Carl VanGessel², Matthew Rouse³, David S. Marshall⁴ and Gina L Brown-Guedira⁵, (1)Department of Crop & Soil Science, Raleigh, NC, (2)Colorado State University, Fort Collins, CO, (3)Cereal Disease Laboratory, USDA-ARS, Saint Paul, MN, (4)Department of Plant Pathology-USDA-ARS, Raleigh, NC, (5)USDA-ARS Plant Science Research, Raleigh, NC

Abstract

Stem rust (*Sr*), caused by *Puccinia graminis* f.sp. *tritici* (*Pgt*), is among the most common and aggressive pathogens causing economic losses in wheat (*Triticum aestivum* L.) production worldwide. This study fine mapped a major QTL in a population of 256 recombinant inbred lines (RILs; hereafter referred to as the LA-population) derived from a cross between the susceptible cultivar LA95135 and the resistant cultivar AGS2000 located on the distal end of chromosome 6DS where at least five other stem rust QTL have been identified (VanGessel 2018; Kassa et al., 2016; Hiebert et al., 2016). We screened the LA-population in growth chambers in the Cereal Disease Laboratory, St. Paul, MN with *Pgt* race TTKSK. Infection type (IT) was scored on a scale of 0 – 4 (0=resistant and 4=susceptible). Resistant RILs had IT scores from 11⁻ to 33 whereas the susceptible lines showed ITs 3⁺3⁺ to 4. The progeny segregated in a ratio of 1 resistant:1 susceptible ($\chi^2_{1,1}=2.83$, 1 d.f., $P = 0.0920$). We exploited exome capture data to discover new polymorphisms for marker saturation in a 10 Mbp region flanking the *Sr* locus. We were able to identify 134 annotated genes and 20 KASP assays were based on two SNP every 1 Mbp used to genotype the LA-population. Based on the genotypic and phenotypic results, we identified a cluster of five candidate NBS-LRR class disease resistance genes located near the reported *Sr* locus. Multiple SNP markers that co-segregated with the resistance allele in the LA-population are being tested for marker-assisted selection (MAS) applications.

PO0943: Wheat, Barley, Oat, and related

Mapping Wheat Stripe Rust Resistance in Overlay × Overland Population

Wardah K. Mustahsan, Kansas State University, Manhattan, KS, Mary J. Guttieri, USDA-ARS, Kansas State University, Manhattan, KS, Robert L. Bowden, USDA-ARS, Department of Plant Pathology Kansas State University, Manhattan, KS, Kimberly Garland-Campbell, Wheat Health, Genetics, and Quality Research USDA-ARS, Pullman, WA and Guorong Zhang, Agricultural Research Center, Hays, KS

Stripe rust, also known as yellow rust, (causal organism *Puccinia striiformis* f. sp. *tritici*) is a major disease hindering wheat production and global food security. Recent race changes have reduced the effectiveness of host plant resistance genes and increased the importance of polygenic resistance. Understanding the genetic mechanisms of disease resistance can aid in producing durable wheat varieties for producers. Stripe rust resistance was evaluated in a hard winter wheat recombinant inbred mapping population, Overlay × Overland. Overlay carries the *T. ventricosum* 2NS segment that contains *Yr17*. Both Overlay and Overland had functionally effective field stripe rust resistance until pathogen race changes occurred in 2010-12. Disease infection type (IT) and severity (SEV) were evaluated in four field trials: Rossville, KS (2018 and 2019), Hays, KS (2019), and Pullman, WA (2019). Environment affected both infection type (IT) and infection severity (SEV). Transgressive segregation was observed

in the population. A genetic linkage map was constructed from 1241 SNPs derived from reduced representation sequencing. The map consisted of 21 linkage groups (LOD = 9.5 r_{max} = 0.3) with a total size of 2221 cM. A KASP marker for the 2NS segment was properly positioned on linkage group 2A. Preliminary QTL analysis identified loci on chromosomes 2A, 2B and 2D associated with IT and SEV using multi-environment averages. The KASP marker for the 2NS segment conferring *Yr17* had an additive effect of -11% on SEV. Resistance alleles on 2A and 2B were contributed by Overley; alleles on 2D were contributed by Overland. Informative SNP sequences will be converted to KASP assays to support breeding efforts for stripe rust resistance in hard winter wheat germplasm.

Submitter's E-mail Address: wkm786@ksu.edu

PE0944: Wheat, Barley, Oat, and related

Genetic Dissection of Stripe Rust Resistance in a Tunisian Common Wheat Landrace Aus26670

Deepak K Baranwal, Harbans Bariana and Urmil Bansal, University of Sydney, Cobbitty, NSW, Australia

The deployment of combinations of resistance genes in future wheat cultivars can save huge wheat production losses resulted by stripe rust (caused by *Puccinia striiformis* f. sp. *tritici*; Pst) epidemics worldwide. This relies on the availability of genetically diverse sources of resistance which in turn is underpinned by continuous discovery. A Tunisian landrace, Aus26670, displayed high level of stripe rust resistance against Australian Pst pathotypes. Aus26670 was crossed with a susceptible genotype Avocet S and genetic analysis based on Aus26670/Avocet S F3 families and F7 RILs (123) indicated the presence of a single all stage resistance (ASR) gene and three adult plant rust resistance (APR) genes in Aus26670. The entire RIL population was genotyped using targeted genotyping-by-sequencing (tGBS) assay. The seedling stripe rust response data were converted to A for resistant and B for susceptible and incorporated into the map. This locus mapped on the long arm of chromosome 2B and was demonstrated to be *Yr72*. QTL analysis suggested the involvement of chromosome 1B, 5A and 7B in controlling APR. The QTL on chromosome 1B corresponded to *Yr29*, whereas the other two QTL appear to be new. F3 families carrying only the chromosomes 5A and 7B linked resistance alleles have been identified and are being advanced. The linked tGBS tags are being blasted against the International Wheat Genome Sequencing Consortium RefSeq v1.0 to design kompetitive allele specific PCR primers for detailed mapping QTL regions on chromosomes 5A and 7B.

PO0945: Wheat, Barley, Oat, and related

Using Forward and Reverse Genetics to Clone the Septoria Nodorum Blotch Susceptibility Gene *Snn5* in Wheat

Katherine L.D. Running¹, Sapna Sharma¹, Steven S. Xu², Phillip McClean¹, Timothy L. Friesen² and Justin D. Faris², (1)North Dakota State University, Fargo, ND, (2)USDA-ARS, Fargo, ND

Parastagonospora nodorum is a necrotrophic fungal pathogen that causes Septoria nodorum blotch (SNB), a disease that reduces yield by decreasing the photosynthetic area of the plant. Wheat lines containing the *Snn5* gene are sensitive to the *P. nodorum*-produced necrotrophic effector (NE) SnTox5, and a compatible *Snn5*-SnTox5 interaction plays a significant role in the development of SNB. The *Snn5* locus was previously mapped to the long arm of chromosome 4B using a doubled haploid (DH) population derived from the *T. turgidum* ssp. *carthlicum* accession PI 94749 and the durum variety Lebsock. Here, we used the same population to conduct saturation mapping and delineate *Snn5* to 2.8 cM, which corresponded to a 1.38 Mb interval containing 18 high-confidence genes in the Chinese Spring reference genome. Chinese Spring and Cadenza were both sensitive to SnTox5 and therefore carry functional alleles of *Snn5*. Both lines also harbored two genes within the *Snn5* candidate region that both had similarity to the SnTox3 sensitivity gene *Snn3-D1*, which was cloned in a parallel project. Markers for the two genes cosegregated with *Snn5*, and analysis of Cadenza TILLING mutants for both genes revealed that one of the genes, but not the other, conditioned SnTox5 sensitivity. Comparative sequence analysis of additional EMS-induced SnTox5-insensitive mutants in the line LP29, which is a SnTox5 sensitive line from the DH population,

further confirmed the candidate gene as *Snn5*. Work is ongoing to further characterize the structure, function, and origin of the *Snn5* gene.

PE0946: Wheat, Barley, Oat, and related

Identification of Molecular Markers for *Septoria tritici* Blotch Resistance in a ‘Madsen x Foote’ RIL Population

David Cobertera, Oregon State University, Corvallis, OR

Septoria tritici blotch (STB) is a foliar disease of wheat (*Triticum aestivum* L.) caused by *Zymoseptoria tritici*. STB is polycyclic disease and represents a major threat to wheat production. High disease pressure can reduce the economic value of the crop by decreasing yield and grain quality. The control of STB relies primarily on fungicides. *Z. tritici* populations have evolved high levels of resistance to certain classes of fungicides, reducing their efficacy. Increasing STB resistance through plant breeding is the most cost-effective way to control this disease. To study STB resistance in wheat, a recombinant inbred line population was developed from a cross between ‘Madsen’ and ‘Foote’ soft white winter wheat. ‘Foote’ (PI 599663) has provided moderate resistance to STB in the Pacific Northwest (PNW), while ‘Madsen’ (PI 511673) is moderately susceptible to STB. The RIL population, consisting of 216 lines, was phenotyped across multiple environments for STB response, and genotyped using Illumina HiSeq 3000 sequencing. The STACKS program was used to select SNPs that were then mapped using MadMapper, SMOOTH, and RECORD. Analysis of variance showed significant differences among phenotypes and genotypes in the RIL population ($p < 0.01$). The best linear unbiased prediction (BLUP) value for each accession across different environments for the severity of STB was used for QTL mapping. Results of quantitative trait loci/locus (QTL) analysis indicated minor genes in 2BL, 3AL, 3BL, 5AL, 6DL, and 7DS. These QTLs could be used to develop molecular markers for genotypic selection of wheat cultivars with improved STB resistance.

PO0947: Wheat, Barley, Oat, and related

Genetic Mapping of Quantitative Trait Loci for *Septoria* Resistance in Winter Wheat Multi-Parent Populations

Adnan Riaz¹, Petra Kockappelgren², Beatrice Corsi³, Fergus Meade⁴, Jie Kang⁵, Nicholas Fradgley³, Stephen Byrne⁶, James Cockram³, Richard Mott⁷, Dan Milbourne⁸ and Ewen Mullins⁹, (1)Teagasc – the Agriculture and Food Development Authority, Carlow, Ireland, (2)Teagasc, Carlow, Ireland, (3)NIAB, Cambridge, United Kingdom, (4)Teagasc, Crop Science Department, Carlow, Carlow, Ireland, (5)AG Research, Dunedin, New Zealand, (6)Teagasc, Crop Science Department, Carlow, Ireland, (7)UCL, London, United Kingdom, (8)Teagasc, Crop Science Department, Oak Park, Ireland, (9)Teagasc-Agriculture and Food Development Authority, Carlow, Ireland

Zymoseptoria tritici is the causative fungal pathogen of *Septoria tritici* blotch (STB) disease of wheat (*Triticum aestivum* L.) that continuously threatens Ireland and Europe’s wheat crop. Under favourable conditions, STB can cause up to 50% yield losses if left untreated. STB is mostly controlled by applying fungicides; however, this incurs an economic loss of more than €1bn annually to the EU. Also, the *Z. tritici* population is developing fungicide resistance, in addition to the increased restriction on fungicide use in the EU; thus, fewer active substances are available for farmers.

Deployment of resistant varieties provides a more sustainable disease management strategy. However, there are no varieties currently on the market that offer an adequate level of resistance against STB. Therefore, innovative breeding methodologies such as marker-assisted selection are needed to develop new varieties with superior resistance.

In this study, we aimed to identify QTL for *Stb* resistance in an 8-way Magic Elite and 16-way MAGIC wheat populations (termed ‘NIAB Diverse MAGIC’). The 8-way MAGIC population, comprising of 720 recombinant inbred lines (RIL), was screened for septoria response in the field under natural infection from 2016 to 2019 for four seasons. Using genotyping information from a 90K single nucleotide polymorphism (SNP) array, we identified quantitative trait loci (QTL) underpinning septoria resistance. Also, screened the 16-way MAGIC population, comprising of >600 RIL, at the seedling stage and adult plant stage in the controlled environment while currently

subjected to multi-location field screening. Using the 35K (SNP) genotyping data, detected a QTL on chromosome 5B, providing resistance to STB at the seedling stage. The genomic regions identified and linked SNPs serve as useful markers for *Stb* resistance, enabling rapid introgression into future bread wheat cultivars.

PE0948: Wheat, Barley, Oat, and related

QTL Mapping of Tan Spot Resistance in Common Wheat PI 277012

Zhaohui Liu¹, Chenggen Chu², Gongjun Shi¹, Hongmei Gong¹, Justin D. Faris³ and Steven S. Xu³,
(1)North Dakota State University, Fargo, ND, (2)Texas A&M AgriLife Research, Amarillo, TX,
(3)USDA-ARS, Fargo, ND

Tan spot, caused by the necrotrophic fungus *Pyrenophora tritici-repentis*, is a common disease of bread wheat and durum wheat worldwide. The disease can cause significant yield and quality losses on susceptible cultivars under favorable conditions. Resistance to tan spot in wheat is complex involving not only the absence of host genes conferring sensitivity to specific necrotrophic effectors (NEs) produced by the fungus but also the presence of QTL conferring race-nonspecific resistance. The common wheat accession PI 277012 possesses high levels of resistance to tan spot. In this work, we conducted genetic analysis of tan spot resistance in PI 277012 using a doubled haploid (DH) population derived from its cross with the highly susceptible spring wheat cultivar 'Grandin'. The DH population was genotyped using 339 SSRs and 2,884 SNPs from the Illumina 9K Infinium array, and phenotyped for sensitivity to Ptr ToxA and reaction to four races of the pathogen at the seedling stage as well as for plant height of the seedling plants. As expected, sensitivity to Ptr ToxA mapped to the known location of the *Tsn1* locus on the long arm of chromosome 5B. The *Tsn1* locus defined a susceptibility QTL for races 1 and 2, which are known to produce Ptr ToxA. A genomic region on chromosome 4B harboring the dwarf gene *Rht-B1* was significantly associated with the plant height at seedling stage and also with tan spot resistance for all the races tested. Other minor QTL were identified on chromosomes 2D, 5D and 7D each of which were associated with specific races. This work indicates that resistance in PI 277012 is largely due to the lack of NE sensitivity genes. The concurrence of resistance and plant height suggests that the hormone homeostasis has a profound effect on tan spot disease development in wheat.

PO0949: Wheat, Barley, Oat, and related

Identification and Sequence Characterization of Recombination Sites Within the Powdery Mildew Resistance Locus *Qpm.tut-4A*

Maciej Majka^{1,2}, Eva Janáková², Irena Jakobson³, Jaroslav Doležal⁴, Kadri Järve³ and Miroslav Valárik²,
(1)Department of Genomics, Cereal Genomics Team of Institute of Plant Genetics PAS Poznan, Poznan, Poland, (2)Institute of Experimental Botany, Olomouc, Czech Republic, (3)Department of Chemistry and Biotechnology, Tallinn University of Technology, Tallinn, Estonia, (4)Institute of Experimental Botany, Centre of the Region Haná for Biotechnological and Agricultural Research, Olomouc, Czech Republic

Crossing of *T. militinae* with the bread wheat cv. Tähti resulted in the introgression line 8.1 showing improved resistance to powdery mildew disease. The major resistance locus *Qpm.tut-4A* was located on long arm of the 4A chromosome within the *T. militinae* introgressed region. The *Qpm.tut-4A* was originally mapped in region flanked by markers *wmc232* and *gwm160* and corresponds to about 10 cM. In an effort to identify the gene, three mapping populations were created. No recombination in the region was observed in the mapping population created from the cross of cv. Tahti and line 8.1. On the other hand, in the mapping population derived from the cross of cv. Chinese Spring and the line 8.1, 32 recombinations enabling reduction of *Qpm.tut-4A* to a 0.22 cM were detected in 7200 plants tested. In the mapping population from the cross of the 8.1 line and cv. Chinese Spring carrying recessive *ph1* locus, recombination rate was elevated 33-fold and 159 recombinant lines were identified. Saturation of regions with recombination events has so far yielded in seven regions with a single recombination event and a size below 3 kbp. Additionally, a 1.2 kbp region with three recombination events was found. This work was supported by the Czech Science Foundation (award 18-11688S), the Czech Republic Ministry of Agriculture (award QK1710302), ERDF project "Plants as a tool for sustainable global development" (No. CZ.02.1.01/0.0/0.0/16_019/0000827), Estonian Ministry of Rural Affairs, State Programme on Plant Breeding, the National Science Center (ETIUDA6-2018/28/T/NZ9/00073), and by the Foundation for Polish Science (FNP).

PE0950: Wheat, Barley, Oat, and related

Wheat Small RNA Response to Powdery Mildew Disease

Manuel Poretti, University of Zürich, Department of Plant and Microbial Biology, Zürich, Switzerland

In the last decade small RNAs have been shown to play a key role in host-microbe interactions. In 2013, the study of Weiberg *et al.* had a major impact on what was known about cross-kingdom exchanges of small RNAs and provided the first example of fungal secreted small RNAs that were capable of silencing the plant's defense genes. These findings open a new scenario where closely living species might use RNA interference (RNAi) as an additional layer of interaction. In the case of a host-pathogen system, this would represent an "extension" of the more classic and wide studied Effector- (ETI) and pathogen-associated molecular pattern (PAMP)-triggered immunity (PTI) models. DNA transposon domestication into cis-regulatory and microRNA (miRNA) sequences is proposed to contribute to abiotic/biotic stress adaptation in plants. The wheat genome is derived at 85% from transposable elements (TEs), and contains thousands of transposons, whose sequences are particularly prone for domestication into miRNA precursors. Therefore, we believe that such a unique and durable host-pathogen relationship could drive the evolution of a TE-based RNAi mechanism that would complement or even regulate the already known ETI and PTI defense mechanisms. In this study, we uncover the existence of wheat-derived miRNAs that can be expressed in response to powdery mildew infection in order to regulate important defense genes in the host. Finally, we show that high-copy transposons contribute to the wheat small RNA immune response to the lineage-specific powdery mildew pathogen.

PO0951: Wheat, Barley, Oat, and related

***Xanthomonas translucens* Commandeers the Host Rate-Limiting Step in ABA Biosynthesis for Disease Susceptibility**

Zhao Peng¹, Ying Hu², Junli Zhang¹, Jose C. Huguet-Tapia¹, Anna Block³, Sunghun Park², Suraj Sapkota⁴, Zhaohui Liu⁴, Sanzhen Liu² and Frank F. White¹, (1)University of Florida, Gainesville, FL, (2)Kansas State University, Manhattan, KS, (3)U.S. Department of Agriculture-Agricultural Research Service, Gainesville, FL, (4)North Dakota State University, Fargo, ND

Plants are vulnerable to disease through pathogen manipulation of phytohormone levels, which otherwise regulate development, abiotic and biotic responses. Here, we show the wheat pathogen *X. translucens* pv. *undulosa* elevates expression of the host gene encoding nine-cis-epoxycarotenoid dioxygenase (*TaNCED-5BS*), which catalyzes the rate-limiting step in the biosynthesis of the phytohormone abscisic acid (ABA) and a component of a major abiotic stress response pathway, to promote disease susceptibility. Gene induction is mediated by a type III transcription activator-like (TAL) effector. The induction of *TaNCED-5BS* results in elevated ABA levels, reduced host transpiration and water loss, enhanced spread of bacteria in infected leaves, and decreased expression of the central defense gene *TaNPR1*. The results represent a novel appropriation of host physiology by a bacterial virulence effector.

PE0952: Wheat, Barley, Oat, and related

Harnessing the Wild Side to Search for Novel Wheat Blast Resistance Genes

Paula Silva¹, Giovana Cruppe², Lidia Calderon³, Narinder Singh⁴, John Raupp², Barbara Valent² and Jesse Poland², (1)Interdepartamental Genetics, Kansas State University, Manhattan, KS, (2)Department of Plant Pathology, Kansas State University, Manhattan, KS, (3)Asociación de Productores de Oleaginosas y Trigo (ANAPO), Santa Cruz de la Sierra, Bolivia (Plurinational State of), (4)Bayer CropScience, Kansas City, MO

When genetic diversity is scarce, plant breeding programs can turn to crop wild relatives as donors of novel sources of diversity. *Aegilops tauschii* is the donor of the D genome of the cultivated bread wheat and has been used as a valuable source of novel disease resistance genes. Wheat blast (WB), caused by *Magnaporthe oryzae* Triticum (MoT) pathotype, is an emerging disease in South Asia that has the potential to devastate wheat production. The objective of this study was to characterize a panel of *Ae. tauschii* accessions for resistance to WB and to identify genomic regions associated with resistance that can be used in marker-assisted selection. We tested 138 accessions of *Ae. tauschii* spp. *strangulata* under controlled conditions in Bolivia. *Ae. tauschii* spikes were inoculated with the MoT Bolivian isolate 008, and disease severity was assessed at 10, 12, 14, 16, and 18 days after inoculation. Genomic regions associated with WB resistance were mapped in a genome-wide association study (GWAS) using the area under the disease progress curve (audpc) as the trait and 13,135 SNP markers. Phenotypic values for audpc ranged from 10 to 790, and six accessions were observed to be more resistant than the highly resistant check. GWAS resulted in the identification of six significant SNPs on chromosomes 4, 2, and 1. This study identified novel genomic regions involved in resistance to WB in a wheat wild relative that have the potential to improve WB resistance in cultivated wheat. Further analysis of these regions will identify candidate genes for these associations.

PO0953: Wheat, Barley, Oat, and related

Phenotypic and Molecular Characterization of Hessian Fly Resistance in Diploid Wheat, *Aegilops tauschii*

Subhashree Subramanyam, USDA-ARS, West Lafayette, IN

The gall midge Hessian fly (Hf, *Mayetiola destructor*; family Cecidomyiidae), is a devastating pest of hexaploid wheat (*Triticum aestivum*) causing significant yield losses. Despite identification and characterization of Hf-responsive defense genes and associated biological pathways against Hf, their functional validation has been challenging due to large genome size, polyploidy, repetitive DNA, and limited genetic resources in hexaploid wheat. The diploid progenitor *Aegilops tauschii*, D-genome donor of modern-day hexaploid wheat, offers an ideal surrogate eliminating the need to target all three homeologous chromosomes (A, B and D) individually making the functional validation of candidate Hf-responsive genes viable. Furthermore, the well-annotated sequence of *Ae. tauschii* genome plus availability of genetic resources amenable to manipulations makes the functional assays less tedious and time-consuming. However, prior to utilization of this diploid genome for downstream studies, it is imperative to characterize its phenotypic and molecular responses to Hf. We screened five *Ae. tauschii* lines and identified two that exhibited a homozygous resistance response to feeding to two Hf biotypes (*L* and *vH13*). The resistant diploid wheat accessions resembled hexaploid wheat in their phenotypic (larval developmental stages, leaf and plant growth, cell wall permeability) and molecular (transcript quantification of Hf-responsive biomarker genes) responses to Hf. Resembling the resistant hexaploid wheat, the diploid accessions mount an early defense strategy involving defense proteins including lectins, secondary metabolites and reactive oxygen species radicals. Our results reveal the suitability of the diploid progenitor for use as an ideal tool for functional genomics research in deciphering the wheat-Hf molecular interactions.

PE0954: Wheat, Barley, Oat, and related

Map-Based Cloning of H34, a Hessian Fly Resistance Gene in Wheat

Nida Ghori, Kansas State University, Manhattan, KS and Mingshun Chen², Yunfeng Xu¹, Zhenqi Su¹, Guihua Bai^{1,2*}

Wheat is a major staple food crop worldwide and various insect pests cause significant grain yield losses to the crop every year. Among them, Hessian fly (*Mayetiola destructor*, HF) is a major destructive pest that can significantly reduce the wheat yield. To date, 35 HF resistance genes have been named. However, lack of diagnostic markers for these genes hampers their deployment in wheat breeding. *H34* is HF resistance gene that was previously mapped on the short arm of chromosome 6B of a wheat cultivar Clark using an F₁₂ recombinant inbred line (RIL) population of

Ning7840 x Clark. Further mapping using 90K Wheat SNP chips identified markers to flank *H34* in a region of 5.3 cM. A cross was made between two RILs, RIL115-S and RIL118-R contrasting for *H34* alleles, and 286 F3 lines were generated to identify heterogeneous inbred families (HIFs). Fine mapping using the HIFs and SNPs developed from 90K SNP chips delimited *H34* in a 4.5 Mb interval. Genotype-by-sequencing (GBS) analysis of the four pairs of near-isogenic lines (NILs) from the selected HIFs identified additional SNPs in the *H34* region and these SNPs were converted into Kompetitive Allele Specific PCR (KASP) markers that further narrowed the *H34* region to 670 kb. The tightly linked flanking markers can be used in marker-assisted selection of *H34* in wheat breeding programs. Also, the successful fine mapping of *H34* to 670 kb interval has laid the foundation for cloning of the gene.

PO0955: Wheat, Barley, Oat, and related

Single Nucleotide Polymorphism Tightly Linked to a Hessian Fly Resistance Gene on Wheat Chromosome Arm 1AS

Yunfeng Xu¹, Fuan Niu¹, Xuming Liu², Lanfei Zhao¹, Mingshun Chen^{2,3} and Guihua Bai^{1,3},
(1)Department of Agronomy, Kansas State University, Manhattan, KS, (2)Department of Entomology, Kansas State University, Manhattan, KS, (3)Hard Winter Wheat Genetics Research Unit, USDA-ARS, Manhattan, KS

Hessian fly [HF; *Mayetiola destructor* (Say)] is one of the most destructive insects of wheat and can cause severe losses in wheat grain yield and quality. Many genes for HF resistance have been reported, but diagnostic markers for those genes are not available for breeding application. To develop user-friendly markers for marker-assisted selection of HF resistance gene *h4* in Java, two populations of recombinant inbred lines (RILs) were developed: 'Java' x 'Bobwhite' with 124 RILs and 'Java' x 'Overley' with 205 RILs. The populations were phenotyped for HF resistance in the greenhouses at Kansas State University using HF Great Plains (GP) biotype. High-density single-nucleotide polymorphism (SNP) linkage maps were developed using single nucleotide polymorphisms (SNPs) generated by genotyping-by-sequencing (GBS). Using the maps, we located *h4* at a ~5.0 cM interval on the distal end of chromosome arm 1AS, which explained 54.9 to 65.0% of the phenotypic variation for HF resistance in the two populations and repeated experiments. Two flanking SNPs, SNP1218-1 and SNP1871-1, were converted to Kompetitive Allele-Specific PCR (KASP) markers for marker-assisted selection. Heterogeneous inbred families (HIFs) were constructed from these RILs that segregated at the *h4* locus; and recombinant near-isogenic lines (NILs) were identified after screening the HIFs using the two flanking markers. Association analysis on a diversity panel of 192 US winter wheat accessions demonstrated that the two flanking KASP markers were tightly associated with *h4* and are useful for selecting *h4* in breeding.

PE0956: Wheat, Barley, Oat, and related

Novel Hessian Fly Resistance QTLs for Hessian Fly Resistance in Wheat Breeding Line SD06165

Lanfei Zhao^{1,2}, Nader R. Abdelsalam^{1,3}, Yunfeng Xu¹, Mingshun Chen¹, Lingrang Kong² and Guihua Bai^{1,4}, (1)Department of Agronomy, Kansas State University, (2)College of Agronomy, Shandong Agriculture University, China, (3)Agricultural Botany Department, Alexandria University, Egypt, (4)USDA-ARS, Manhattan, KS

Hessian fly (HF), *Mayetiola destructor* (Say), is one of the most destructive pests in wheat (*Triticum aestivum* L.) worldwide. Growing resistant cultivars is the most effective approach to minimize Hessian fly damage. To identify quantitative trait loci (QTLs) for HF resistance, a recombinant inbred line (RIL) population was developed by crossing HF resistant wheat line SD06165 to a susceptible line OK05312, genotyped with 1,709 single nucleotide polymorphisms (SNPs) generated from genotyping-by-sequencing (GBS), and phenotyped for HF resistance in greenhouses. Two novel QTLs for HF resistance were identified from SD06165. The major QTL, designated as *H35*, was closely linked to SNP marker SDOKSNP7679 on chromosome 3BS that explained 23.8% and 36.0% of the phenotypic variations, and the minor QTL, designated as *H36*, was flanked by SNP markers SDOKSNP1618 and SDOKSNP8089 on chromosome 7AS and explained 8.5% and 13.1% of the phenotypic variation in two experiments. Significant interaction was detected between the two QTLs. Seventeen SNPs that tightly link to *H35* and eight SNPs that tightly link to *H36* were converted to Kompetitive Allele Specific Polymerase (KASP) markers for selecting these QTLs in breeding programs.

PO0957: Wheat, Barley, Oat, and related

Genetic Diversity and Assessment of Wheat Genotypes Under Water Stress Condition

Nadia M. Khan, University of Karachi, Pakistan, Karachi, Pakistan

Wheat is regarded as one of the most important worldwide cereal crop and utilized as staple food in most of the countries including Pakistan. Wheat productivity restricted by several biotic and abiotic stress. Among all the stresses, water stress is one of the major environmental stress that limits wheat productivity worldwide. Experiment was conducted in field to assess the influence of water stress on growth and yield attributes of wheat genotypes. Moreover, stem water soluble carbohydrates (SWSC) and total protein contents were also estimated. Water stress was applied after flowering stage. Morphological traits were recorded under irrigated and non-irrigated condition. The results suggested that genotypes not only performed differently but also responded variably to water stress, hence some being more tolerant than others. Pakistani wheat genotypes Mehran-89, NR-234 and Saleem-2000 by and large proved to be more stress tolerant as compared to other genotypes. Stem water soluble carbohydrates are a source of carbon for grain filling in wheat that become more important in maintaining yield of grain during post-anthesis period when photosynthesis decline due to water stress. Results showed that increased SWSC in Tatara, Zamindar-80 and LU-26 under water stress. Moreover, total protein contents were also increase in these genotypes. These trait could therefore be considered potential indicators for indirect selection of wheat varieties with water stress tolerance, this approach might lead to new genotypes with high yield stability and potential that in turn will result in superior performance in water-stress environment.

PE0958: Wheat, Barley, Oat, and related

Novel Quantitative Trait Loci for Grain Cadmium Concentration in Common Wheat (*Triticum aestivum* L.)

Yusuke Ban¹, Goro Ishikawa², Hiroaki Ueda³, Naoyuki Ishikawa¹, Keita Kato¹, Kanenori Takata⁴, Minoru Matsuyama⁵, Hirokazu Handa⁶, Toshiki Nakamura⁷ and Mikiko Yanaka⁸, (1)NARO Western Region Agricultural Research Center, Fukuyama, Japan, (2)NARO Institute of Crop Science, Tsukuba, Japan, (3)Ehime Research Institute of Agriculture, Forestry and Fisheries, Matsuyama, Japan, (4)Obihiro University of Agriculture and Veterinary Medicine, Obihiro, Japan, (5)Hyogo Prefectural Technology Center for Agriculture, Forestry and Fisheries, Kasai, Japan, (6)Kyoto Prefectural University, Kyoto, Japan, (7)NARO Tohoku Agricultural Research Center, Morioka, Japan, (8)NARO Kyushu Okinawa Agricultural Research Center, Chikugo, Japan

Cadmium (Cd) is as an extremely toxic metal that can contaminate agricultural soils. To reduce the risk of Cd intake in food cereals, the development of cultivars with low grain Cd concentration (GCC) is an effective countermeasure. We analyzed quantitative trait loci (QTLs) for GCC in common wheat (*Triticum aestivum* L.) in a doubled haploid (DH) population developed from ‘Chugoku 165’ (low GCC) × ‘Chukei 10-22’ (high GCC). We found novel loci for GCC on the short arm of chromosome 4B and on the long arm of chromosome 6B. These QTLs accounted for 9.4%–25.4% (4B) and 9.0%–17.8% (6B) of the phenotypic variance in the DH population. An association analysis with 43 cultivars identified 3 loci at these QTLs: *QCdc.4B-kita*, *QCdc.6B-kita1*, and *QCdc.6B-kita2*. In contrast to durum wheat and barley, no QTL was detected on the chromosomes of homeologous group 5 for *heavy metal PIB-type ATPase 3*. Thus, Cd accumulation in grain of common wheat might involve another mechanism. These results will contribute to marker-assisted selection for low GCC in breeding of common wheat.

PO0959: Wheat, Barley, Oat, and related

Variation in Phosphorus and Sulfur Content Shapes the Genetic Architecture and Phenotypic Associations within the Wheat Grain Ionome

Andrii Fatiukha¹, Valentyna Klymiuk¹, Zvi Peleg², Yehoshua Saranga², Ismail Cakmak³, Tamar Krugman⁴, Abraham Korol¹ and **Tzion Fahima**¹, (1)Institute of Evolution and the Department of Evolutionary and Environmental Biology, Faculty of Natural Sciences, University of Haifa, Haifa, Israel, (2)R. H. Smith Institute of Plant Science & Genetics in Agriculture, The Hebrew University of Jerusalem,

Rehovot, Israel, (3)Faculty of Engineering & Natural Sciences, Sabanci University, Istanbul, Turkey, Istanbul, Turkey, (4)Institute of Evolution, University of Haifa, Haifa, Israel

Dissection of the genetic basis of wheat ionome is crucial for understanding the physiological and biochemical processes underlying mineral accumulation in seeds, as well as for efficient crop breeding. Most of the elements essential for plants are metals stored in seeds as chelate complexes with phytic acid or sulfur-containing compounds. We assume that the involvement of phosphorus and sulfur in metal chelation is the reason for strong phenotypic correlations within ionome. Adjustment of element concentrations for the effect of variation in phosphorus and sulfur seed content resulted in drastic change in phenotypic correlations between the elements. The genetic architecture of wheat grain ionome was characterized by quantitative trait loci (QTL) analysis using a cross between durum and wild emmer wheat. QTL analysis of the adjusted traits and two-trait analysis of the initial traits paired with either P or S considerably improved QTL detection power and accuracy, resulting in the identification of 105 QTLs and 617 QTL effects for 11 elements. Candidate gene search revealed some potential functional associations between QTLs and corresponding genes within their intervals. Thus, we have shown that accounting for variation in P and S is crucial for understanding of the physiological and genetic regulation of mineral composition of wheat grain ionome and can be implemented for other plants. In addition, the identified wild emmer wheat QTL alleles leading to increase in GPC grain protein content and essential elements, such as Zn and Fe, emphasize the potential of wild relatives for improvement of nutrition quality of crops.

PE0960: Wheat, Barley, Oat, and related

Genetic Mapping of Phosphorus Use in Hard Red Spring Wheat Cultivated in the Northern Great Plains Over the Last Century

Jason Fiedler, USDA Agricultural Research Service, Fargo, ND, Naa Korkoi Ardayfio, USDA-ARS, Fargo, ND, Jessica Paler, NDSU Dept. of Soil Science, Fargo, ND and Jay R Goos, Dept. of Soil Science, North Dakota State University, Fargo, ND

Hard red spring wheat is a high-protein bread wheat and a major crop in North Dakota. We identified approximately 150,000 genome-wide single nucleotide polymorphism (SNP) markers in cultivars grown throughout the region. Sub-structure analysis identified several clusters of the lines, with release year and breeding program origin mostly influencing the structure designations. Diversity analysis showed that that allelic richness decreased over time as new cultivars became increasingly more similar to each other than in previous time-frames. The population was subjected to growth under varied phosphorus-limiting conditions, and agronomic trait data was collected to identify cultivars with high phosphorus-use efficiency (PUE). Genome-wide association analysis identified several SNPs that were associated with the traits. The globe is running out of rock phosphate ore needed to make P fertilizers and the efficient use of this element will become important to maintain yield potential in the future. The markers identified here and with subsequent fine-mapping have potential for tracking these traits in wheat breeding programs to maintain high PUE and mitigate future risk.

PO0961: Wheat, Barley, Oat, and related

GrainGenes: New Content and Browsers for the Small Grains Community

Victoria Carollo Blake¹, Gerard R. Lazo², Steve L. Michel³, Eric Yao⁴, Halise Busra Cagirici⁵, Laurel Cooper⁶, Charlene P. Wight⁷, Nicholas A. Tinker⁷ and Taner Z. Sen⁵, (1)Montana State University, Bozeman, MT, (2)USDA ARS WRRRC, Albany, CA, (3)USDA-ARS-WRRRC / GrainGenes, Albany, CA, (4)University of California, Berkeley, CA, (5)USDA-ARS, Albany, CA, (6)Dept. of Botany and Plant Pathology, Oregon State University, Corvallis, OR, (7)Agriculture and Agri-Food Canada, Ottawa, ON, Canada

GrainGenes (graingenes.org; wheat.pw.usda.gov) is the centralized, curated USDA-ARS database for wheat, barley, oat, and rye, ensuring long-term data sustainability for small grains researchers. The bulk of the curated data in GrainGenes resides on a MySQL database, where in 2019 curators added ~2100 QTL from the durum cv. Svevo sequencing project, 197 QTL listed in the Catalog of Gene Symbols for Wheat (WGC), and 60 map sets from legacy genetic maps for oat in collaboration with colleagues at Agriculture and Agri-Food Canada. In the summer of 2019, GrainGenes launched a project to represent all genes in the WGC as interactive gene records in the database.

GrainGenes hosts JBrowse genome browsers that are shared with The Triticeae Toolbox (T3) database and in 2019 added assemblies for the durum cv. Svevo RefSeq 1.0, wild emmer wheat cv. Zavitan 2.0, and a track for the 1000 wheat exomes SNP data sets were aligned to the IWGSC wheat Chinese Spring RefSeq 1.0 assembly. GrainGenes is an active participant of the Wheat Information System (WheatIS; wheatis.org) and in 2019 indexed QTL, germplasm, maps and genes from the Wheat Gene Catalogue for WheatIS.

PE0962: Wheat, Barley, Oat, and related

Bart 2.0: A PacBio Iso-Seq Based Reference Transcript Dataset for Barley

Max Coulter¹, Ronja Wonneberger², Runxuan Zhang³, Micha Bayer⁴, Nicola McCallum⁴, Nils Stein⁵, John Brown¹ and Robbie Waugh³, (1)University of Dundee, Dundee, United Kingdom, (2)Leibniz Institute of Plant Genetics and Crop Plant Research, Gatersleben, Stadt Seeland, Germany, (3)The James Hutton Institute, Dundee, United Kingdom, (4)The James Hutton Institute, United Kingdom, (5)Leibniz Institute of Plant Genetics and Crop Plant Research, Gatersleben, Germany

How plants respond to internal and external cues at the transcriptional level are important questions that could be answered using RNA-Seq analysis. Phenotypic responses reflect both changes in gene expression and alternative splicing. Salmon/Kallisto generate rapid and accurate transcript quantification from RNA-seq data but require an accurate and comprehensive reference transcript dataset (RTD). RTDs have been developed for Arabidopsis and barley but are largely based on short read sequencing with its inherent difficulties in accurate assembly of whole transcripts.

Here, we present progress on developing a new RTD for barley using cv. Barke, a European 2-row spring variety, and a combination of long read Pacbio Iso-Seq and short read Illumina RNA-Seq data from RNA libraries made from 20 different tissues and treatments. Iso-Seq data was analysed with Iso-seq3 and TAMA. Long read sequences have high error rates and a major source of transcript sequence errors are misalignments due to errors close to splice junctions. We have developed a method to 1) use the highly accurate splice junction data from short reads to identify and remove transcripts with false splice junctions and 2) rationalise transcript start and end sites and generate high confidence transcripts. A parallel pipeline using multiple assemblers generates transcripts from short reads to supplement transcripts and genes not covered by Iso-Seq reads.

The resulting RTD will be used to analyse RNA-Seq data from 200 European spring barley varieties to identify associations between transcript abundance and yield traits and construct gene co-expression networks associated with yield characteristics.

PO0963: Wheat, Barley, Oat, and related

The Maltome: Description and Analysis of the Transcriptome from Malting Barley

Marcus A Vinje, USDA-ARS, Cereal Crops Research Unit, Madison, WI, Khoa Le, University of Minnesota, Department of Plant Pathology, St. Paul, MN and Cory D. Hirsch, University of Minnesota, St. Paul, MN

Barley malt is the primary ingredient used to brew beer. Simply explained, brewing beer is accomplished in three steps: malting, mashing, and fermentation. Malting is a form of controlled germination that occurs when raw barley is imbibed to ~45 % moisture, incubated at cool temperatures with rotation for 4-5 days, and kilned to stop the biological processes. During malting and mashing, large polymers (e.g. starch and protein) are converted into smaller polymers (e.g. maltose and amino acids) that are utilized by yeast during the fermentation process. The malting process has been honed over millennia from steeping barley in flowing rivers to large industrial floor malting facilities and over this time malting quality parameters have been established that barley and malt must meet in order to make quality. However, a complete understanding of the genes/transcripts and, most importantly, gene networks involved in the malting process have not been firmly established. Therefore, this work was undertaken to 1.) establish a model of the malting transcriptome (i.e. The Maltome) that describes the regulation of genes and their networks 2.) identify genetic networks and genes/gene families that have the potential to influence specific malting quality parameters 3.) identify putative malting quality genes from previous research in this data set and 4.) predict genes that putatively could influence malting quality parameters based on known mechanisms. We have malted in

triplicate an elite malting barley cultivar, Conrad, grown under multiple environments at the USDA, Cereal Crops Research Unit, which is the facility and malting parameters used by most public U.S. barley breeders. Seven time points (Dry, 0 Days of Germination [DoG], 1-5 DoG) were sampled for each of the three reps for a total of 21 Illumina libraries. All 21 libraries were sequenced on the Illumina HiSeq2500 sequencer on three 1 x 100 lanes.

PE0964: Wheat, Barley, Oat, and related

Genetic Dissection of the Semi-Dwarfing Genes *sdw1* and *ari-e* in a Barley MAGIC Population

Viet Dang¹, Camilla Beate Hill¹, Xiao-Qi Zhang¹, Tefera Tolera Angessa¹, Lee-Anne McFawn² and Chengdao Li^{1,2}, (1)Western Crop Genetics Alliance, Agricultural Sciences, College of Science, Health, Engineering and Education, Murdoch University, Perth, WA, Australia, (2)Department of Primary Industries and Regional Development, Perth, WA, Australia

The major semi-dwarf genes *sdw1* and *ari-e* have been widely utilised in many barley breeding programs worldwide. Semi-dwarf cultivars are characteristic for their reduced plant height, improved stress resistance, higher grain yield and quality compared to tall varieties. In this study, we identified a novel *sdw1* allele in cv. Lockyer and investigated the effect of *sdw1* and *ari-e* and their interactions on plant height, flowering time and grain yield using a Multi-parent Advanced Generation Inter-cross (MAGIC) population generated from four elite barley cultivars. Allele-specific markers combined with Kompetitive Allele Specific Polymerase Chain Reaction (KASP) assays were used for fast and accurate screening of plants with different allele combinations. Results showed that the novel *sdw1* allele from cv. Lockyer showed similar effects on plant growth and phenology compared to the well-known *sdw1.d* allele. The two semi-dwarf genes *sdw1* and *ari-e* showed an additive effect for plant height with a height reduction up to 27 cm. The *ari-e* gene was found to cause earlier flowering from 3 to 6 days and reduced the delay in flowering time caused by the *sdw1* gene. The two semi-dwarf genes showed a significant boost in grain yield of 20% in the *sdw1* plants and up to 28% in plants with the two combined semi-dwarf genes. The results presented are useful resources to develop high yielding and better adapted barley cultivars

PO0965: Wheat, Barley, Oat, and related

Genebank Genomics - a Barley Case Study

Nils Stein, Leibniz Institute of Plant Genetics and Crop Plant Research, Gatersleben, Germany

Plant Genetic Resources (Plant GenRes) may hold the key for adapting crops to a changing climate. Their actual use in crop improvement, however, is limited and is in stark contrast to their potential and promise. Major international initiatives like DivSeek have advocated for systematically generating genotypic and phenotypic data for genetic resources stored in genebanks. At IPK we have accomplished a pilot study on the genotyping of all >20,000 barley accessions of our ex situ gene bank. This study has opened new avenues in the way we can make information available to users of genetic resources, the way we (want to) manage genetic resources and it has revealed the composition of global domesticated barley diversity – laying the foundation for exploring the pan-genome of the species for the profit of research and application. By the barley example, IPK has initiated in the context of other national or international collaborative projects the systematic genotyping of further collections such as of wheat, *Solanaceae* crops, *Phaseolus* beans thus paving the way of turning the gene bank from a seed archive into a bio-digital resources center.

PE0966: Wheat, Barley, Oat, and related

Introgression from Cultivated Barley in a Wild Barley Collection

Chaochih Liu¹, Li Lei², Mingqin Shao³, Connor Depies³, Malik Samuel³, Jerome D. Franckowiak³, Brian J. Steffenson⁴ and Peter L. Morrell³, (1)Department of Plant and Microbial Biology, University of Minnesota, Saint Paul, MN, (2)DOE Joint Genome Institute, Berkeley, CA, (3)Department of Agronomy and Plant Genetics, University of Minnesota, Saint Paul, MN, (4)Department of Plant Pathology, University of Minnesota, Saint Paul, MN

A number of studies have reported genotypic evidence that phenotypically wild accessions of wild barley (*Hordeum ssp. spontaneum*) have been subject to introgression from cultivated barley. We used 318 total accessions from the

Wild Barley Diversity Collection (WBDC), which is part of the extensive collections of wild barley accessions that are maintained for their utility in crop improvement. With comparable genotype and exome capture data, we performed identity by state comparisons between the 318 WBDC accessions and cultivated barley to identify potential regions of genomic introgression. We found multiple WBDC accessions show evidence of introgression. Using the genomic intervals for well-characterized genes involved in domestication and improvement, we examined evidence for introgression at genomic regions potentially important for maintaining a wild phenotype. Assessing the size of runs of identity by state suggests that most, but not all, introgression has occurred very recently.

PO0967: Wheat, Barley, Oat, and related

Genetic Analysis of Spikelet Abortion in Barley

Roop Kamal¹, Quddoes H. Muqaddasi¹, Thirulogachandar Venkatasubbu¹ and Thorsten Schnurbusch^{1,2}, (1)Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), OT Gatersleben, Stadt Seeland, Germany, (2)Faculty of Natural Sciences III, Institute of Agricultural and Nutritional Sciences, Martin Luther University Halle-Wittenberg, Halle (Saale), Germany

The potential to increase barley grain yield lies in the indeterminate nature of its Inflorescence Meristem (IM). The IM produces spikelets, the basic reproductive units of grasses, which are inevitably linked to the grain yield. However, spikelet abortion – a phenomenon in which spikelets abort after a specific developmental stage, imposes a bottleneck on increasing grain yield potential. Although several hypotheses explaining this phenomenon are documented, the exact underlying genetic and molecular mechanism is still unclear. In this study, we examined Potential Spikelet Number at maximum yield potential (PSN^{MYP}) and Final Spikelet Number at heading (FSN^{Heading}) stages and computed the Spikelet Abortion (SA) in order to better understand the phenomenon of spikelet abortion genetically. We, therefore, used the traits mentioned above in a panel of 432 six-rowed spring barley accessions representing worldwide genetic diversity for a genome-wide association study (GWAS). All accessions were evaluated in replicated field trials. Phenotypic data analyses showed significant genotypic variation for all traits with high broad-sense heritability estimates (0.71 – 0.89). The Potential Spikelet Number at maximum yield potential has a strong Pearson's correlation of 0.65 with spikelet abortion. All accessions were fingerprinted with genotyping-by-sequencing that yielded 14,913 high-quality markers. Our GWAS identified Quantitative Trait Loci (QTL) for PSN^{MYP} on chromosome 7H, for PSN^{Heading} on chromosome 4H and for SA on chromosome 7H. Our preliminary results shed light on the quantitative genetic nature of spikelet abortion by studying component traits. This serves as a source of marker-assisted breeding, and also provide a resource for future gene identification.

PE0968: Wheat, Barley, Oat, and related

Identification of Yield-Related Genes in Barley Using a Multi-Layered Genomics and Transcriptomics Approach

Ronja Wonneberger¹, Allison Haaning², Max Coulter³, Wenbin Guo⁴, Robbie Waugh⁵, Gary J. Muehlbauer² and Nils Stein⁶, (1)Leibniz Institute of Plant Genetics and Crop Plant Research, Gatersleben, Stadt Seeland, Germany, (2)Department of Agronomy and Plant Genetics, University of Minnesota, St. Paul, MN, (3)The University of Dundee, United Kingdom, (4)University of Dundee, Dundee, United Kingdom, (5)The James Hutton Institute, Dundee, United Kingdom, (6)Leibniz Institute of Plant Genetics and Crop Plant Research, Gatersleben, Germany

Increasing crop yield through breeding requires a thorough understanding of the genetic architecture of several morphological and physiological traits, as well as of the interaction and regulation of genes affecting these traits. In this project, we aim to identify and characterize genes regulating yield-related traits in a panel of 211 two-rowed spring barley cultivars representing the European breeding germplasm of the last century. The panel was phenotyped for more than 15 traits, mainly grain and spike traits, in Gatersleben (Germany), Dundee (Scotland) and St. Paul, MN (USA) in 2019.

Commonly, the search for genetic variation focuses on genome-wide association studies (GWAS) to find sequence variation associated with a trait, while gene expression variation as a driver of phenotype is often neglected. Transcriptome-wide association studies (TWAS) complement GWAS by allowing the identification of associations

between transcript abundance variation and the trait of interest. It also allows us to identify candidate genes, especially in genomic regions with high linkage disequilibrium.

High-resolution GWAS will be performed using SNPs obtained from RNAseq and whole genome sequencing. The creation of a transcriptome atlas of six tissues will allow TWAS to elucidate how gene expression influences key traits in barley and how these genes are regulated throughout plant development. Transcript abundance will be accurately quantified using the Pacbio long read-based reference transcriptome BaRT 2.0 currently under development. Gene co-expression networks will shed light on how key regulatory genes interact to determine yield. Ultimately, candidate genes will be further validated in appropriate mutant populations.

PO0969: Wheat, Barley, Oat, and related

Dissecting the Genetic Basis of Spikelet Survival Rate in Barley

Yongyu Huang, Leibniz Institute of Plant Genetics and Crop Plant Research, Gatersleben, Germany

Selection for inflorescence architecture with more grain production and yield is common to many cereal crops such as wheat and barley. In barley, the number of grains depends on the number of fertile spikelets initiated along its inflorescence axis, and the spikelet reaches its maximum number at a certain growth stage, i.e. at around awn primordium stage. After this, the number of spikelet will be gradually decline in a basipetal pattern due to spikelet abortion, leading to a substantial amount of spikelet loss. Therefore, lifting spikelet survival rate represents an intriguing approach to enhance grain yield in barley. In this study, we combined quantitative trait locus mapping, genome-wide association studies and mutated gene analysis to study the genetic basis of this pre-anthesis spikelet loss.

PE0970: Wheat, Barley, Oat, and related

Genetic Mapping of Floret Number per Spikelet in Barley (*Hordeum vulgare* L.)

Guojing Jiang, IPK, Gatersleben, Germany

The inflorescences of the *Triticeae* tribe, which includes wheat (*Triticum* sp. L.) and barley (*Hordeum vulgare* L.), are made up of a group of specialized short branches known as spikelets. The spikelet axis called rachilla has the ability to differentiate continuously to develop into one to multiple florets. In barley, the diminutive rachilla shows vestigial meristematic activity, leading to a single-flowered spikelet. Spikelet triplets distichously attached to each rachis node give barley a well-regulated row-type spike. Although several genes regulating spike development (majorly row-type) in barley have been identified, the genetic mechanisms of regulation for inflorescence structure are still poorly understood. Here, we report an inflorescence mutant *multiflorus2.b* (*mul2.b*) in barley, which shows multiple florets in lateral spikelets. The *mul2* mutants fail to suppress the determinacy of the rachilla in lateral spikelets, producing up to three florets in lateral spikelets. Interestingly, the rachillae of central spikelets remain determinate producing single florets, indicating rachilla determinacy between central and lateral spikelets may have separate regulatory patterns. Genetic mapping using an F₃ segregating population enabled us to position *mul2* on chromosome 2H to a 10.94 cM region. We propose that the *mul2* locus is involved in regulation of rachilla development.

PO0971: Wheat, Barley, Oat, and related

Exploring the Realm of Possibilities: Trying to Predict Promising Crosses through Genomic Mating in Barley

Amina Abed and Francois Belzile, Laval University, Quebec, QC, Canada

Genomic prediction approaches aim to increase the rate of genetic gain for complex quantitative traits. Introducing genomic prediction at the cross-selection stage (genomic mating) is much more palatable to established breeding programs than at progeny-selection stages (genomic selection). In this study, we tried to assess the reliability of a predictive model in an applied breeding context. We developed a genomic prediction model to examine a large set of simulated progenies obtained from crosses made between lines in the training population. We predicted the mean (μ) and genetic variance (V_G) for each cross. In addition, we wanted to identify promising crosses in which the

strong unfavorable correlation between the most important traits in barley breeding, grain yield (GYD) and deoxynivalenol (DON) accumulation, was predicted to be weakened or abolished. Whether for GYD or DON, the pattern of μ and V_G was similar as the majority of crosses were predicted to generate populations with variable but limited and fewer exhibiting moderate and large. To assess the reliability of the predicted value (μ and V_G) of a cross, we used a retrospective approach whereby we followed the persistence of crosses through the course of the breeding process. Crosses predicted to be superior produced progeny that persisted longer in the breeding program. This suggests that the predictions made regarding the performance of the simulated crosses are not too far from reality. The predicted correlations between traits known to be correlated (e.g. DON-GYD) were concordant with observed and expected correlations suggesting that the properties of these simulated progeny are in agreement with what is expected for these traits. Overall, of the 30,000 possible crosses that could potentially be made between lines comprising the training population, only 2.2% were predicted to show a low correlation between GYD and DON and just 0.13% were predicted to produce progeny in which the top lines could combine high GYD with improved DON. Even in the absence of empirical proof that genomic prediction can outperform classical practice, the results obtained here appear encouraging with regard to the potential of such an approach in barley breeding programs.

PE0972: Wheat, Barley, Oat, and related

Multi-Trait Genomic Prediction Model for Agronomic and Malting Quality Traits in Barley (*Hordeum vulgare* L.)

Madhav Bhatta¹, Lucia Gutierrez¹ and Ariel Castro², (1)Department of Agronomy, University of Wisconsin - Madison, Madison, WI, (2)College of Agriculture - UDELAR, Paysandu, Uruguay

Plant breeders regularly evaluate multiple traits across multiple environments, which opens an avenue for using multiple traits in genomic prediction models. Our goal was to compare the genomic predictive ability of single and multi-trait prediction models with two cross-validation schemes (CV1, predicting new lines with genotypic information only and CV2, predicting partially phenotyped lines using both, genotypic and phenotypic information from correlated traits) for agronomic and malting quality traits in barley (*Hordeum vulgare* L.). The predictive ability was similar for single and multi-trait models to predict new lines. However, the predictive ability for agronomic traits was considerably increased when partially phenotyped lines were used. Similarly, a considerable increase in the predictive ability of malting quality traits was observed when correlated traits were used. This study showed the potential of improving genomic prediction of complex traits by incorporating the information from multiple traits (cost-friendly and easy to measure traits) collected throughout breeding programs which could assist in speeding up breeding cycles.

PO0973: Wheat, Barley, Oat, and related

Evaluation of Cell Wall Biosynthetic Genes in Barley (*Hordeum vulgare* L.)

Alicia Nesbary¹, Jamie Sherman² and Christopher Botanga¹, (1)Chicago State University, Chicago, IL, (2)Montana State University, Bozeman, MT

Lodging is a highly undesirable attribute for cereal crop plants because of increased disease infestations, decreased grain quality, uneven maturity, and increased harvesting costs. Lodging in barley, *Hordeum vulgare* can cause a 65% reduction in grain yield. There are variations in straw (stem) strength among barley cultivars. Stiff-strawed cultivars tend to show higher resistance to lodging. The aim of this study is to correlate the relationship between lignin deposition and structural differences in the cell walls of barley internodes at three growth stages: stem elongation, booting, and flowering. Another objective was to identify genes that could be used as markers for straw stiffness. Microscopy revealed that lignin was localized in the sclerenchyma ring in the cortex. It also revealed that stiff-stemmed cultivars have thicker cell walls than weak-stemmed barley cultivars. PCR primers designed from cell wall biosynthetic genes from rice ESTs, showed largely monomorphic PCR products, suggesting that the cell wall genes evaluated are present in both the stiff- and weak-strawed genotypes of barley. This also suggests that the cell wall biosynthetic genes evaluated are likely orthologs and highly conserved among the three species: barley, rice and wheat. Draft sequences of the barley genome were used to amplify entire coding regions of select cell wall genes. This study also revealed that there are significant differences in the expression of genes involved in lignin biosynthesis, with activity mainly in older internodes at the flowering stage. These results elucidate our understanding of cell wall thickness and lignin deposition during barley development. This project was supported in part by the USDA NIFA (Award #: 2016-70003-24775).

PE0974: Wheat, Barley, Oat, and related

Mutations in the HvMKK3 and HvAlaAT1 Genes Affect Barley Pre-Harvest Sprouting and After-Ripened Seed Dormancy

Justin Vetch¹, Jamie Sherman¹, John M. Martin¹, Mike Giroux¹ and **Jason Walling²**, (1)Montana State University, Bozeman, MT, (2)USDA-ARS Cereal Crops Research Unit, Madison, WI

Preharvest sprouting (PHS) is a natural phenomenon that negatively impacts various crops across the globe when late season rainfall causes seeds to germinate prior to harvest. PHS damage could be mitigated by incorporation of genetic alleles that impart desired levels of dormancy at specific times in grain maturity. Toward this goal, 114 barley varieties were assessed for dormancy at physiological maturity and in after-ripened grains. Three genes previously associated with dormancy in barley or wheat (*HvAlaAT1*, *HvMKK3*, and *HvMFT*) were sequenced from all lines and assessed for allelic diversity. The resulting alleles were assessed for association with dormancy at each of the maturity time points. Here we report missense mutations in both *HvAlaAT1* and *HvMKK3* that significantly affect dormancy levels in barley.

PO0975: Wheat, Barley, Oat, and related

The Influence of Allelic Variant Combination on Barley Lodging Tolerance

Toi John Tsilo¹, PE Kyeswa², Tebogo Oliphant¹, Tsepiso Hlongoane³ and Timmy A Baloyi², (1)Agricultural Research Council - Small Grain Institute, Bethlehem, South Africa, (2)Agricultural Research Council, Bethlehem, South Africa, (3)Agricultural Research Council - Small Grains, Bethlehem, South Africa

Barley (*Hordeum vulgare L.*) is important for beer production. Barley lodging is a problem which has a negative influence on yield and quality of barley thereby resulting in loss of profit. Lodging is a complex trait that is affected by various abiotic and biotic factors. Thus, achieving lodging resistance and good malting quality is a major challenge due to multiple gene interactions and genes deployed to improve lodging. Careful selection of the best allele combinations is essential to mitigate the influence dwarfing/lignification genes may have on malting quality and yield. A wide range of barley genotypes have been screened with molecular markers and the different allele combinations reveal the influence selected allele combinations have on lodging tolerance.

PE0976: Wheat, Barley, Oat, and related

Revisiting the Labile (lab) Locus of Barley

Jeyaraman Rajaraman, Independent HEISENBERG Research Group Plant Architecture, Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), OT Gatersleben, Stadt Seeland, Germany, Twan Rutten, Structural Cell Biology, Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), OT Gatersleben, Stadt Seeland, Germany and Thorsten Schnurbusch, Martin Luther University Halle-Wittenberg, Faculty of Natural Sciences III, Halle, Germany; Independent HEISENBERG Research Group Plant Architecture, Leibniz-Institute of Plant Genetics and Crop Plant Research (IPK), OT Gatersleben, Stadt Seeland, Germany

The labile-barley genotype exhibits an irregular number of fertile florets per rachis node compared to that of two- or six-rowed counterparts (1). The factors responsible for this random reduction in florets still remain elusive. Previously, in an effort to identify the locus, we mapped the lab locus on chromosome 5HL using F2 populations (2).

In the current study, we generated a high-density genetic map using homozygous F2:S5 recombinant inbred lines (RILs) combined with genotyping by sequencing (GBS) approach and reevaluated the labile phenotype in a detailed quantitative manner under three different environments. Linkage analysis revealed a new single QTL peak on the long arm of chromosome 2H, consistent with an interval of ~8.4 Mbp, challenging our previous report. Efforts are currently being undertaken to high-resolution mapping of the 2H locus using F2 populations and residually heterozygous lines (RHL) from the RIL population.

Furthermore, 3D image reconstruction of a labile-barley rachis node suggests a developmental implication associated with the floret meristem. To address this, we are also investigating the optimal growth condition for consistent expression of the labile phenotype which will be later used for transcriptome analysis. Taken together, our study aims to solve the longstanding mystery behind the yet unknown genes and the mechanism underlying this random phenomenon of floret elimination/death in labile-barley.

References:

Åberg e, Wiebe GA (1945) Irregular barley, *Hordeum irregulare*, sp.nov. J Wash Acad Sci 35:161–164
Youssef, H.M., R. Koppolu, et.al, (2014) Genetic mapping of the labile (lab) gene: a recessive locus causing irregular spikelet fertility in labile-barley (*Hordeum vulgare labile*). Theor Appl Genet 127: 1123-1131

PO0977: Wheat, Barley, Oat, and related

Identification and Mapping of Novel Seedling Gene Conferring Resistance to *Puccinia hordei* in Barley

Mehnaz Mehnaz¹, Peter Dracatos¹, Robert Park¹, Anh Pham², Timothy March², Andreas Maurer³, Klaus Pillen³, Kerrie Forrest⁴, Tejal Kulkarni⁴ and Davinder Singh¹, (1)University of Sydney, Sydney, NSW, Australia, (2)University of Adelaide, Urrbrae, SA, Australia, (3)Martin-Luther-University, Halle, Germany, (4)Genomics & Cell Sciences | Agriculture Victoria Research, Bundoora, VIC, Australia

Barley leaf rust (BLR), one of the most destructive diseases of barley, is caused by the fungal pathogen *Puccinia hordei*. Genetic resistance is considered to be the most effective, economical and eco-friendly approach to minimize the losses caused by this disease. A study was undertaken to discover and map a new gene conferring seedling resistance to *P. hordei* (RphHEB-04-101) in a *Hordeum spontaneum*-derived barley line HEB-04-101. Genetic analysis of an F3 population derived from a cross between barley line HEB-04-101 (seedling resistant to Australian *P. hordei* pathotype 5457P+) and *Hordeum vulgare* cultivar Flagship (seedling susceptible to 5457P+) confirmed inheritance for a single dominant gene in line HEB04-101. Targeted genotyping by sequencing (tGBS) analysis using 25 homozygous resistant and 25 homozygous susceptible F3 families identified putatively linked markers on the long arm of chromosome 5H spanning a physical interval between 622-669 Mbp based on the 2017 Morex barley reference genome assembly. To date four seedling resistance genes (Rph2, Rph9, Rph12 and Rph25) and one adult plant resistance gene (Rph20) have been mapped on chromosome 5H. The pathotype used in this study (5457P+) is virulent on seedlings of differentials carrying these four genes, demonstrating that the BLR resistance in HEB-04-101 is novel. We developed 56 CAPS (Cleaved amplified polymorphic sequences) markers within the physical region harbouring the gene and genotyped the F3 population (n=125) using 15 polymorphic markers to further refine the map location. Two flanking markers HOR_5H_640.33 and HOR_5H_640.71 were identified at genetic distances of 2.0 and 0.8 cM respectively. Further fine mapping is underway using a high resolution F2 mapping population (n > 1500 individuals) to develop markers tightly linked to the resistance locus and to identify candidate genes for RphHEB-04-101-mediated resistance.

PE0978: Wheat, Barley, Oat, and related

An *Mla* Homolog Confers Susceptibility to Spot Blotch caused by *Bipolaris sorokiniana* in Barley Bowman

Yueqiang Leng¹, Mingxia Zhao¹, Rui Wang¹, István Molnár², Jaroslav Doležel², Burkhard Steuernagel³, Jason D. Fiedler⁴, Brian J. Steffenson⁵ and Shaobin Zhong¹, (1)Department of Plant Pathology, North Dakota State University, Fargo, ND, (2)Institute of Experimental Botany, Centre of the Region Haná for Biotechnological and Agricultural Research, Olomouc, Czech Republic, (3)The Sainsbury Laboratory and John Innes Centre, Norwich, United Kingdom, (4)USDA-ARS, Cereal Crops Research Unit, Edward T. Schafer Agricultural Research Center, Fargo, ND 58102, Fargo, ND, (5)Department of Plant Pathology, University of Minnesota, Saint Paul, MN

In this study, we cloned a gene (*Scs6*) conferring susceptibility to spot blotch caused by *Bipolaris sorokiniana* in barley cv. Bowman using the “MutChromSeq” approach. *Scs6* was previously fine-mapped in a genomic region harboring the *Mla* locus on chromosome 1H (Leng et al. 2018). Further map-based cloning of the gene was hindered

due to genome complexity and low recombination rate in the target region. An ethyl methane sulfonate (EMS) mutant population was generated in cv. Bowman and 14 resistant mutants were identified. Flow-sorted 1H chromosomes of Bowman (wild type) and five of the EMS mutants were sequenced and compared, leading to identification of a candidate gene encoding a coiled-coil nucleotide binding site Leucine-rich repeat (NLR) protein, which was deleted in one mutant and altered in four other different mutants. PCR-amplification and Sanger-sequencing confirmed all five mutations and identified additional two mutants with mutations in the same candidate gene. This gene is nearly identical to one *Mla* allele cloned from wild barley *Hordeum vulgare* subsp. *spontaneum* with only four nucleotide differences. It had 99.17%, 99.10%, 92.59% and 92.34% identity to *Mla16-1*, *Mla18-1*, *Mla25-1* and *Mla38-1*, respectively, the only four *Mla* genes coding unusually polymorphic CC domain and lacking detectable resistance activity to *Blumeria graminis* f. sp. *hordae* among 23 *Mla* genes studied (Seeholzer et al. 2010). Virus-induced gene silencing confirmed this *Mla* homolog is *Scs6* with dominant susceptibility function. Sequence analysis indicated *Scs6* is highly conserved among susceptible barley accessions but it was mutated or deleted in resistant barley genotypes.

PO0979: Wheat, Barley, Oat, and related

High-Resolution Mapping of the Barley Mild Mosaic Virus (BaMMV) Resistance Gene *rym15*

Yaping Wang¹, Antje Habekuss², Dragan Perovic² and Frank Ordon², (1)Julius Kühn-Institut (JKI) Federal Research Centre for Cultivated Plants, Institute for Resistance Research and Stress Tolerance, Quedlinburg, Germany, (2)JKI, Julius Kuehn Institute, Quedlinburg, Germany

Barley is the second most important cereal crop in Europe. *Barley mild mosaic virus* (BaMMV) which is transmitted by the soil-borne protist *Polymyxa graminis* has a serious impact on barley yield. In previous studies on doubled haploid (DH) lines of the cross ‘Chikurin Ibaraki 1’ (R) × ‘Plaisant’ (S), the BaMMV resistance gene *rym15* was mapped on the short arm of chromosome 6H flanked by markers at a distance of 6.4 cM. The present work aims to construct a high-resolution mapping population of *rym15*, narrow down the target region and saturate the map with the final aim to isolate *rym15*.

Two crosses derived from the resistant barley cv. ‘Chikurin Ibaraki 1’ and susceptible cultivars ‘Igri’ and ‘Uschi’ were used for the construction of a medium-resolution mapping populations of *rym15*. Segregation ratios of 342 and 180 F₂ plants of the ‘Igri’ × ‘Chikurin Ibaraki 1’ and ‘Chikurin Ibaraki 1’ × ‘Uschi’, i.e. 251(S) : 91(R) and 139(S) : 41(R), respectively, fit to a ratio of 3s:1r ($\chi^2=0.472$, $\chi^2=0.474$), suggesting the presence of one recessive resistance gene. Six published SSR markers and 8 KASP markers developed based on the 50K Illumina array data were used for the medium-resolution mapping. Genetic maps were constructed, and new robust co-dominant flanking markers were identified. Furthermore, in order to construct the high-resolution mapping population of *rym15*, 168 and 286 F₂ plants recombinant in the respective interval were identified by screening 2174 and 5724 F₂ plants from the respective crosses. A set of 212 SNPs between the flanking markers was identified by comparing the 50K Infinium Illumina array and GBS data of susceptible and resistant bulks. These markers will be used for marker saturation of the target locus and narrowing down the candidate gene interval with the final aim to facilitate the positional cloning of *rym15*.

PE0980: Wheat, Barley, Oat, and related

Improvement of Virus Resistance Breeding in Barley by the Help of *H. bulbosum*

Behnaz Soleimani¹, Julia Brandes¹, Dragan Perovic¹, Antje Habekuss¹, Viktor Korzun², Klaus Oldach², Neele Wendler³, Hélène Pidon³, Nils Stein⁴ and Frank Ordon¹, (1)JKI, Julius Kuehn Institute, Quedlinburg, Germany, (2)KWS LOCHOW GmbH, Bergen, Germany, (3)Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Seeland, Germany, (4)Leibniz Institute of Plant Genetics and Crop Plant Research, Gatersleben, Germany

Hordeum bulbosum, the only member in the secondary gene pool of *Hordeum vulgare*, holds resistances and tolerances against various pathogens, for example against *Barley mild mosaic virus/Barley yellow mosaic virus* (BaMMV/BaYMV) or *Barley yellow dwarf virus* (BYDV). Both diseases cause high yield losses in barley. Furthermore, the control of the aphid-transmitted BYDV is becoming difficult due to governmental regulations concerning insecticides. The use of chemicals to control BaMMV/BaYMV, transferred by the soil-borne protist

Polymyxa graminis, is not possible. Thus, breeding for resistance is the only possibility to protect barley against these diseases.

Different *H. bulbosum* introgression lines carrying resistance against BaMMV/BaYMV (*Rym16^{Hb}*) and tolerance against BYDV (*Ryd203S11^{Hb}*) on chromosome 2HL were characterized. The sizes of the introgression fragments were calculated based on the barley reference sequence and resulted in a size of 4.2 Mb for the locus *Ryd203S11^{Hb}* and 3 Mb for the locus *Rym16^{Hb}*.

The analysis of 10,000 F₂ plants carrying *Ryd203S11^{Hb}* and 4440 F₂ plants carrying *Rym16^{Hb}* with co-dominant flanking markers resulted in 34 recombinant F₃ plants, which will be used to construct high resolution mapping populations. The recombination rate within the introgression is about 14 times reduced compared to the intraspecific recombination rate within the barley genome, most likely caused by the incomplete homology between the genome of *H. vulgare* and *H. bulbosum*.

As a basis for isolating the respective genes via a map-based cloning approach, recombinant plants will be selfed, phenotyped and saturated with markers using Exome capture, GBS and Illumina 50K data. A non-gridded BAC library will be used to construct a physical map of the target region of *Ryd203S11^{Hb}*. This map will help to identify candidate genes located in the *H. bulbosum* introgression fragment. In addition, the resistance of *Rym16^{Hb}* will be analyzed by using resistance gene enrichment sequencing (RenSeq).

PO0981: Wheat, Barley, Oat, and related

Elucidating Mechanisms of Non-Host Resistance Towards *Septoria passerinii* and *Zymoseptoria tritici* in Wheat and Barley

Cassidy R. Million, USDA-ARS, West Lafayette, IN, Charles F. Crane, USDA-ARS and Department of Botany and Plant Pathology, Purdue University, West Lafayette, IN and Stephen B. Goodwin, USDA ARS, West Lafayette, IN

Non-host resistance, by definition, is complete, durable resistance of a plant species to a potential pathogen. *Septoria passerinii* is a pathogen of barley but not wheat, while *Zymoseptoria tritici* (previously *Mycosphaerella graminicola*) infects wheat but not barley. Wheat displays gene-for-gene resistance responses to *Z. tritici* with ~21 genes mapped to date. Little is known about the mechanisms of resistance and specifically whether different resistance genes in the same host utilize common or variable mechanisms to achieve the defense response. Thus, the objective of this study is to identify the mechanisms of resistance in the host and non-host responses towards two pathogens, *S. passerinii* and *Z. tritici*, and to test whether similar mechanisms are being utilized by the host and non-host responses. To test this, RNA sequencing was carried out at 5 time points ranging from fungal penetration to the late infection stage (1, 3, 6, 10, and 17 days after inoculation). The non-host interaction consisted of *S. passerinii* inoculated onto a wheat genotype that is susceptible to *Z. tritici* (Taichung 29), while *Z. tritici* (isolate IPO323) was inoculated onto a normally susceptible barley genotype. Host interactions consisted of *Z. tritici* inoculated onto two wheat lines containing resistance genes *Stb2* (on wheat chromosome 1BS) and *Stb3* (7AS) plus the Taichung 29 susceptible genotype, and *S. passerinii* inoculated onto resistant and susceptible barley genotypes. Our findings suggest that non-host resistance responses have different timings and gene expression patterns compared to host (R gene) responses. In both interactions at 1 dai, a large number of differentially expressed genes (DEGs) were measured. In the non-host response, a similar number of DEGs are measured throughout the entire infection period, whereas in the host response DEGs decreased over the infection period. The non-host response of barley to *Z. tritici* is induced, not constitutive. This study gives a greater insight into the mechanisms of non-host resistance as well as compatible and incompatible interactions in the host response.

PE0982: Wheat, Barley, Oat, and related

Population Genomics of *Brachypodium hybridum*

Li Lei¹, Sean Gordon², Virginia Markham¹, Pilar Catalan³ and John Vogel¹, (1)DOE Joint Genome Institute, Berkeley, CA, (2)DOE Joint Genome Institute, Walnut Creek, CA, (3)Department of Agriculture (Botany), High Polytechnic School of Huesca, University of Zaragoza, Huesca, Spain

Polyploid genomes are characteristic of grasses being developed as biomass crops and many grain crops. Therefore, a deeper understanding of population genomics and genome evolution in polyploid genomes would be useful for developing improved crop varieties for both food and fuel. Unfortunately, biomass and grain crops are difficult experimental subjects because of their large, complex genomes, outbreeding nature and/or their large physical size. Thus, a simple model system to study polyploid genome regulation and evolutions would be very useful. *Brachypodium hybridum*, is an allotetraploid with subgenomes derived from ancestors similar or identical to the diploid species *B. distachyon* and *B. stacei*. All three species have very compact genomes, small stature and are easily grown and manipulated in the laboratory. We have sequenced and assembled the genomes of the three species. Our previous results suggested that at least two hybridization events happened in the speciation of *B. hybridum* around 1.4 and 0.14 million year ago. However, the details about the number of polyploidization events and subsequent interbreeding are unknown. Therefore, the aim of this study is to explore the population structure, population differentiation, demographic history, and the genetic basis of the climatic adaptation of *B. hybridum* using pan-genome and population genomic approaches. Longer term, we plan to define regulatory elements that can be ultimately be used to improve crops.

PO0983: Wheat, Barley, Oat, and related

Marker Identification Associated With Important Quality and Morphological Traits of NDSU Oat Breeding Lines

Naa Korkoi Ardayfio¹, Sepehr Mohajeri Naraghi², Michael S. McMullen² and Jason D. Fiedler³, (1)USDA-ARS, Fargo, ND, (2)North Dakota State University, Fargo, ND, (3)USDA-ARS, Cereal Crops Research Unit, Edward T. Schafer Agricultural Research Center, Fargo, ND 58102, Fargo, ND

Oat is an important cereal crop with enormous benefits as food for humans and fodder for animals. The nutritional components of oat include high β -glucan, low carbohydrates, high protein content, significant dietary fiber, and vitamins. The North Dakota State University oat breeding program has been developing oat lines for the last century and has released several important cultivars. Conventional breeding methods can be accelerated with molecular techniques that focus on identifying molecular markers to predict traits. Three generations of experimental data were used to detect qualitative and quantitative trait loci (QTL) in NDSU oat breeding lines. Markers that flag these QTL have potential to be used for marker assisted selection to increase the efficiency to develop new lines.

PE0984: Wheat, Barley, Oat, and related

Insights into Rye Biology and Triticeae Relationships Based on a Chromosome-Scale Genome Assembly

Nils Stein, Leibniz Institute of Plant Genetics and Crop Plant Research, Gatersleben, Germany

Rye, *Secale cereale*, belongs to the Triticeae tribe in the *Poaceae*. Opposite to the global importance of its close relatives barley and wheat, rye is economically relevant mainly to the Northern European countries. It is produced as animal feed and biofuel crop but also for human food. Rye biology and life history is in contrast to wheat and barley for many aspects. Rye is a secondary domesticate which traveled to Europe as a weed in barley and wheat fields. It is a self-incompatible and out-crossing species with very good winter hardiness. Heterosis in rye is high and is exploited efficiently in CMS-facilitated hybrid breeding. Its diploid genome is estimated to comprise 7-8 Gigabases which is about 50% larger than the closely related diploid Triticeae genomes e.g. of barley (5 Gbp). An international consortium used now de novo short read sequencing-by-synthesis and assembly combined with high density genetic mapping, Hi-C analysis and optical mapping to produce a high quality chromosome-scale sequence assembly, which greatly facilitates the application of molecular genetic tools and strategies in research and breeding for crop improvement and for reaching a better understanding of the mechanisms underlying rye's distinct biological and genetic features.

PO0985: Wheat, Barley, Oat, and related

Genomic Selection for Forage Yield in Tetraploid Perennial Ryegrass

Stephen Byrne¹, Patrick Conaghan², Sai Krishna Arojju¹ and Dan Milbourne³, (1)Teagasc, Crop Science Department, Carlow, Ireland, (2)Teagasc, Grassland Science Research Department, Carlow, Ireland, (3)Teagasc, Crop Science Department, Oak Park, Ireland

Studies have shown that perennial ryegrass breeders have succeeded in achieving significant breeding gain for forage yield. Indirect selection using genome-wide markers is one tool that offers breeders an opportunity to further accelerate genetic gain for forage yield. In perennial ryegrass this can be achieved by enabling multiple cycles of genomic selection (GS) to be completed in the same time it takes to perform a single cycle of conventional selection. We established a small population of half-sib families by intercrossing plants of an elite commercial tetraploid cultivar. Maternal plants were genotyped using a genotyping-by-sequencing strategy and half-sib progenies were phenotyped for forage yield in replicated sward plots over 2 years under two cutting managements. Predictive models were developed using maternal genotype data and predictive ability for forage yield varied across different cutting periods and managements. Even low to moderate predictive abilities should encourage use of GS in routine selection given our ability to complete at least five cycles of GS in the time it takes to complete a single cycle of conventional selection.

PE0986: Wheat, Barley, Oat, and related

Genome Wide Association Study of Quinoa

Laura Kate Christensen, Brigham Young University, Provo, UT

Quinoa has great agronomic potential as it is high in nutritional value and is well adapted to abiotic stresses that limit yields of other staple crops. However, for quinoa cultivation to become more widespread, certain traits need to be improved. A required step in this process is the identification of genes that affect desired traits. A common method used to identify genes underlying traits of interests is genome wide association studies (GWAS) which is a computational observational study that tests for association of a set of genetic variations with phenotypic variation of different individuals. Through GWAS analysis, possible linkage between observed phenotypes and specific gene can be determined, which can lead to better selection of desirable traits.

In this study, linkage between observed phenotypes and genotypes is tested through a series of data analysis, GWAS analysis, selection of candidate genes and growing of mutant plants to confirm a gene's effect upon a trait of interest. This project has implications to help quinoa to grow outside of its native area, as well as introducing new technology and skills into the lab groups who study quinoa.

PO0987: Wheat, Barley, Oat, and related

Evaluation of Two Cycles of Genomic Selection in an Intermediate Wheatgrass Breeding Program

Jared Crain, Kansas State University, Manhattan, KS, Lee DeHaan, The Land Institute, Salina, KS and Jesse Poland, Department of Plant Pathology, Kansas State University, Manhattan, KS

Perennial grains could provide a host of ecosystem and environmental services, yet large scale adoption of perennial grains require having economically viable crop yields. Intermediate wheatgrass (*Thinopyrum intermedium*) has been undergoing domestication as a perennial grain crop since the mid 1980's. While phenotypic breeding has produced large breeding gains, over 10% genetic gain per cycle, it is estimated that another 20 and 110 years of equal breeding gains will be required to reach the grain yield and seed size of annual wheat, respectively. Beginning in 2017, genomic selection (GS) has been used in The Land Institute's breeding program to overcome the long-estimated times to achieve a comparable product as annual wheat. Each year over 4,500 genotypes have been profiled using genotyping-by-sequencing, with GS used to predict genotype performance. The 100 best genotypes have been moved immediately to the crossing block for intermating allowing one cycle to be completed per year. An additional 1,000 genotypes are planted in the field as the GS training population. Correlation between the seedling predicted genotype performance and the observed field observations have ranged from 0.14 to 0.73. The realized selection differential has ranged from 10-23% superior for the selected parents compared to the random training population. Utilizing the GS pipeline has resulted in reducing cycle time by half which should theoretical double the rate of genetic gains. Our current results indicate that greater than 10% genetic gain per year can be achieved for selected traits using GS, speeding the development of perennial grains.

PE0988: Wheat, Barley, Oat, and related

***SPO11-1* Mutants Provide Insights into Recombination Progression and Crossover Formation in Hexaploid Bread Wheat**

Kim Osman, University of Birmingham, Birmingham, United Kingdom

In most eukaryotes the distribution of meiotic crossovers (COs) along chromosomes is non-random due to multiple levels of control. In wheat and other cereals the predominantly distal location of COs creates a problem of linkage-drag in the recombinationally 'cold' centromere/proximal and interstitial regions where agronomically important traits cannot be readily separated from undesirable ones. We are investigating the factors influencing CO formation in hexaploid wheat, which has 3 sub-genomes (A, B and D), to find ways to modulate the process and unlock genetic diversity for crop improvement. In one approach we used CRISPR-Cas9 editing to introduce targeted mutations in *SPO11-1*, which encodes a key component of the highly conserved DNA topoisomerase VI-like recombination initiation complex. Fortuitously, one round of editing generated different edits in 5 of the 6 gene copies. The single remaining D-genome wild-type (WT) copy was sufficient to retain fertility and enable propagation and crossing to generate a series of lines with different allele combinations. In all cases a single WT copy from any of the sub-genomes was sufficient to maintain fertility. Two triple mutant lines which differ only in their B-genome allele were completely sterile. Cytological analysis of Pollen Mother Cells (PMCs) revealed a complete absence of chiasmata/COs in one of the mutants and an extremely low level of residual chiasmata in the other, suggesting that these lines might be useful hosts for re-targeting recombination in the wheat genome. Immunolocalisation revealed that although the chromosome axis appeared to form normally in both mutants, chromosomes failed to synapse. Interestingly, the weaker of the two mutants still appeared to produce DNA double-strand breaks (DSBs). The nature of these is still being investigated and current data will be presented. However, the observation that DSBs are repaired yet are unable to progress efficiently to COs has clear implications for CO re-targeting experiments and underlines the need for a greater understanding of the biology of recombination initiation in large-genome polyploid plants. Ongoing detailed analysis of the effects of various allele combinations in the wheat *SPO11-1* edited lines will contribute to this.

PO0989: Wheat, Barley, Oat, and related

Field Transcriptomics Identifies a Novel Wheat Susceptibility Factor Which Modulates Resistance to Rust

Pilar Corredor-Moreno, Phoebe E Davey and Diane GO Saunders, John Innes Centre, Norwich, United Kingdom

Wheat yellow rust, caused by the obligate biotrophic fungus *Puccinia striiformis* f. sp. *tritici* (*Pst*), is a major threat to wheat production worldwide and can lead to total crop loss when untreated. Identifying host pathways targeted by the pathogen is crucial to get a better understanding of the host mechanisms involved in the defence response. We used transcriptomics data obtained from *Pst*-infected field samples collected from wheat varieties showing different susceptibilities to *Pst*. Differential expression analysis identified changes in amino acid metabolism that led to the identification of *TabCAT1*, a gene encoding a wheat branch-chain aminotransferase. *TabCAT1* expression peaks at 24 hours post-inoculation in a susceptible interaction with modern *Pst* UK isolates while the expression was reduced in a resistant interaction. To further explore the function of this protein in the defence response, we developed *Tabcat1* tetraploid double mutant lines using the wheat TILLING (Targeting Induced Local Lesions in Genomes) population. *Tabcat1* mutant lines showed a dramatic reduction in susceptibility to *Pst*, suggesting a potential role as a wheat susceptibility factor. Mutant lines exhibited constitutive upregulation of pathogenesis-related (PR) genes in the absence of infection and this observation was replicated in hexaploid wheat when *TabCAT1* was silenced using the barley stripe mosaic virus (BSMV). *Tabcat1* mutant lines also accumulated salicylic acid (SA) in the absence of pathogen infection, indicating that mutant lines have a constitutively activated defence machinery. These results suggest that *TabCAT1* coordinates the activation of SA-mediated defence responses in wheat and could be exploited in resistance breeding to eliminate known susceptibility from commercial cultivars.

PE0990: Other Category

The Genomics Education Alliance (GEA)

Vincent Buonaccorsi, Juniata College, Huntingdon, PA, Anne Rosenwald, Georgetown University, Washington, DC, Rochelle Tractenberg, Georgetown University, Jason Williams, CyVerse, Tucson, AZ and Douglas L. Chalker, Washington University in St Louis, St Louis, MO

The Genomics Education Alliance represents a group of life science educators who have experience engaging students in Course-based Undergraduate Research Experiences (CUREs) in genomics and bioinformatics. Because we are convinced that CUREs are effective for students to learn both key concepts and the practice of science, we have come together to identify and overcome common barriers to put such experiences within the reach of all life science faculty and students. To achieve these goals, the GEA will: 1) host core bioinformatics tools, 2) curate and/or develop curriculum and faculty training resources, and 3) curate CURE assessment materials. We plan to curate a wide variety of freely available materials both from our existing genomics CUREs and new resources we create. The GEA will utilize the cyberinfrastructure provided by CyVerse to ensure sufficient compute capacity for faculty to use GEA resources in the classroom. Ultimately, we aim to facilitate efforts by faculty who build their own genomics CURE using our optimized resources. We are now recruiting faculty to pilot a set of stand-alone lessons meant to support CUREs in three areas: lessons on examining gene sequence similarities using BLAST, understanding eukaryotic gene structure by using a genome browser, and investigating gene expression by using basic tools for RNA-seq analysis. To learn more about these lessons and to sign up for this pilot, please visit the GEA web site at <https://gea.qubeshub.org/lessons>. The GEA is supported by National Science Foundation Research Coordination Network for Undergraduate Biology Education (NSF RCN-UBE) grant #DBI 1827130.

PO0991: Other Category

Developing Ecofabs As a Reproducible System to Study the Root Microbiome

John Vogel¹, Peter Andeer², Kateryna Zhalnina³ and Trent Northen³, (1)DOE Joint Genome Institute, Berkeley, CA, (2)Lawrence Berkeley National Laboratory, (3)Lawrence Berkeley National Laboratory, Berkeley, CA

In recent years the microbiome has been shown to play a major role in plant and animal health and productivity. However, despite its importance, we know little about how plants and microbes interact to shape microbiome composition and function. The lack of reproducible model ecosystems is a limitation to our ability to study and understand the microbiome. Currently, most studies either focus on a single isolate and a single plant growing in a petri dish or use they use 16s or ITS amplicon sequencing to survey microbial abundance in field samples. While the former approach allows researchers to control and manipulate nearly all aspects of the system, the lack of diverse microbes creates a highly artificial system that is often not relevant to agriculture. The latter approach captures much of the complexity of what goes on in the field, but it is impossible to precisely replicate and control the microbial community in the field. Our efforts are aimed at developing a middle ground where we use a small, defined microbial communities in a fully controlled ecosystem. To enable this, we developed small growth containers (EcoFABs) that allow us to non-destructively monitor root and microbial growth, sample the growth medium for metabolomics all while avoiding contamination from environmental microbes. Progress towards increasing throughput using mass production and robotic handling will be presented.

PE0992: Methods: Markers

Holistic Genotyping of Amplicon Panels with Spades and Blastn

Charles F. Crane, USDA-ARS and Department of Botany and Plant Pathology, Purdue University, West Lafayette, IN

Amplicon panels fall into three categories, the first two of which have been used widely as genetic markers: microsatellites, RADseq/ddRADseq, and targeted panels based on exons or transposon insertion junctions. Traditional genotyping of these panels has rested on detection and scoring of a single measure of variation, such as microsatellite length or SNP base within the amplicon. The focus on a single measure has limited the amount of allelic variation that can be reliably scored at a single locus. Here I discuss another, more holistic approach, in which sequenced amplicons are assembled with SPAdes and then scored on the basis of blastn bitscore. This tactic allows various combinations of individual SNPs and/or repeat-count variants to be scored as distinct alleles. Simulation results indicate that distinct alleles can be recognized with as little as 1% variation among them, given typical base-calling error frequencies in Illumina reads. Application of the method to a set of eight population samples of Hessian

fly (*Mayetiola destructor* (Say)) confirmed that the method was robust even though up to half of the reads within any individual fly did not map to the primary alleles because of sequencing errors. Problems with the method include very uneven sampling among competing amplicons in multiplexed PCR, allele-specific differences in amplification rate, difficulty proving that a given amplicon arises from one locus, and stutter if the amplicons contain a dinucleotide-based microsatellite. The ability to recognize n alleles in a heterozygous n -ploid parent should greatly facilitate linkage and deletion mapping in polyploid species, since the allele-phasing problem mostly vanishes and allele dosage can be estimated from sufficient depth of read coverage.

PO0993: Wheat, Barley, Oat, and related

Efficient Characterization of Tetraploid Wheat Plant Genetic Resources for Wheat Resilience Improvement

Marco Maccaferri¹, Elisabetta Mazzucotelli², Giuseppe Sciarra³, Raj Pasam⁴, Anna-Maria Mastrangelo⁵, Danara Ormanbekova¹, Martina Bruschi³, Elisabetta Frascaroli¹, Silvio Salvi¹, Simona Corneti³, Sandra Stefanelli¹, Reem Joukhadar⁶, Francesca Desiderio², Steven Xu⁷, Matthew Hayden⁸, Hakan Ozkan⁹, Benjamin Kilian¹⁰, Aldo Ceriotti¹¹, Curtis Pozniak¹², Karim Ammar¹³, Hans Brown¹⁴, Luigi Cattivelli¹⁵, Filippo Maria Bassi¹⁶ and Roberto Tuberosa¹⁷, (1)Department of Agricultural and Food Sciences (DISTAL), University of Bologna, Bologna, Bologna, Italy, (2)CREA-Research Centre for Genomics and Bioinformatics, Italy, (3)Department of Agricultural and Food Sciences (DISTAL), University of Bologna, Bologna, Italy, (4)Agriculture Victoria, AgriBio, Centre for AgriBioscience, Melbourne, Australia, (5)CREA Research centre for cereal and industrial crop, Foggia, Italy, (6)DEDJTR, AgriBio, Bundoora, Australia, (7)USDA-ARS, Edward T. Schafer Agricultural Research Center, Fargo, ND, (8)Agriculture Victoria, AgriBio, Centre for AgriBioscience, Melbourne, Vic, Australia, (9)University of Cukurova, Adana, Turkey, (10)The Global Crop Diversity Trust (GDCT), Bonn, Germany, (11)National Research Council (CNR), Institute of Agricultural Biology and Biotechnology (IBBA), Italy, (12)Crop Development Centre, University of Saskatchewan, Saskatoon, SK, Canada, (13)CIMMYT, Mexico D.F., Mexico, (14)CIMMYT, International Maize and Wheat Improvement Center, Mexico, (15)CREA - Research Centre for Genomics and Bioinformatics, Fiorenzuola d' Arda, Italy, (16)ICARDA, Rabat, Morocco, (17)Department of Agricultural and Food Sciences, University of Bologna, Bologna, Italy, Bologna, Italy

Marco Maccaferri, Department of Agricultural and Food Sciences (DISTAL), University of Bologna, Bologna, Italy

Climate change poses major challenge for global wheat production and thus food security. Drought, heat, salinity and recurrent rust, septoria and fusarium epidemic waves are among the most common abiotic and biotic stresses limiting wheat crop productivity and sustainability worldwide. If adequately characterized, natural variants present in underutilized plant genetic resources (PGRs) are potentially valuable for providing improvements on resilience to abiotic stress and durable resistance to plant pathogens. In this context, Golden-standard reference genomes as well as characterized PGRs are instrumental. Both *Triticum aestivum* and *Triticum turgidum* ssp. *durum* share BB and AA genomes inherited from wild and domesticated emmer throughout complex steps of domestication, spread and migration events from the Fertile Crescent to diverse environments and agro-ecological conditions. With the aim of facilitating germplasm characterization and use, we assembled two comprehensive and complementary tetraploid collections: *i*) the Global Durum Wheat Panel (GDP) and the Tetraploid wheat Global Collection (TGC, Maccaferri et al. *Nature Genetics* 2019). GDP was established through a cooperative effort in the frame of the Wheat Initiative. GDP, currently maintained by ICARDA, was assembled by bringing together the durum wheat cultivated germplasm from more than 50 countries worldwide, including ca. 500 cultivars and 400 landraces, pre-breeding lines and emmer. The Tetraploid wheat Global Collection (TGC), of 1,856 single-seed descent derived-genotypes, represents 11 tetraploid BBAA wheat taxa covering the whole distribution range. The Illumina 90K wheat SNP array was used to characterize both collections. Ca. 20,000 unique, non-redundant, single Mendelian SNP markers that were both genetically and physically mapped were used to obtain a haplo-based map of germplasm. We provide a detailed dissection of the huge reservoir of genetic diversity available in these two tetraploid genetic resources and examples of successful utilization of these resource to conduct GWAS for traits related to improvement of wheat sustainability and adaptation to climate change effects.

PE0994: Natural Populations

Genome Assembly Variation and Time-Series Association in Brachypodium Species

Jared Streich¹, Jonathan Romero², Ashley Cliff², David Kainer¹ and Daniel Jacobson¹, (1)Oak Ridge National Laboratory, Oak Ridge, TN, (2)University of Tennessee/Oak Ridge National Laboratory, Oak Ridge, TN

Brachypodium distachyon and other Brachypodium species have been ideal model grass organisms, but currently lack a mitochondrial genome. Plant mitochondrial genomes are challenging to assemble due to a long inverted repeat. We used the closely related and fully assembled wheat mitochondrial genome *Triticum aestivum* as a backdrop for creating the first *B. sylvaticum* MTgenome using PacBio long-reads. First we assembled BWA mappable reads into contigs using the SPADes Assembler creating the first intermediate Brachypodium MTgenome. Contigs from this assembly were then randomly broken down into fastq files of varying length from 100-300bp and used to call SNP and indel variants on the *Triticum aestivum* genome. Variants were then imputed onto the wheat MTgenome using GATK's 'FastAlternateReferenceMaker' function. After several rounds of polishing through contig assembly, contigs to short reads, variant calling, and variant imputation, a high quality Brachypodium *sylvaticum* MTgenome assembly was created. Gene models were then run using Prokka to annotate the MTgenome. A similar process was used to create MTgenomes for several accessions of other Brachypodium species including 121 *B. distachyon*, 2 *B. stacei*, and 2 *B. hybridum* using the *B. sylvaticum* MTgenome. The final assemblies were then used for trans-chromosomal linkage using Custom Correlation Coefficient (CCC), environmental association (GWATS), and demography analysis.

PO0995: Maize, Sorghum, Millet, Sugar Cane, and related

Towards a Viable Perennial Grain Sorghum through Integrated Genetic and Economic Analysis

Tara Maireid Conway, Plant Genome Mapping Laboratory, UGA, Athens, GA, Wenqian Kong, University of Georgia, Athens, GA, T. Stan Cox, The Land Institute, Salina, KS and Andrew H. Paterson, Plant Genome Mapping Laboratory, University of Georgia, Athens, GA

The replacement of annual cereal crops with perennial varieties has great potential for increased ecosystem resilience and improved food security. Successful establishment of such cropping systems requires development of both system "hardware" (perennial grain varieties) and "software" (agronomic and ecological systems facilitated by the perennial germplasm). Integration of the "hardware" and "software" development holds great potential for increasing the efficiency of perennial grain system establishment. Sorghum (*S. bicolor*) the world's fifth most important cereal crop, is an ideal target for perenniality due to its close relation to two wild perennial grasses, *S. halepense* and *S. propinquum*, and its potential to grow across highly diverse landscapes. The key to perenniality lies in the development of rhizomes, subterranean stems that sprout to form the next season's crop. Quantitative trait locus (QTL) analysis of a novel F2:3 population derived from an *S. bicolor* x *S. halepense* cross illuminated genomic regions pertinent to perenniality and can contribute to marker-assisted selection in breeding. Simultaneous investigations into Georgia sorghum farm budgets highlighted differential yield targets contingent on perennial cropping scenario as well as other breeding targets to further improve viability for Georgia farmers.

PE0996: Rice

SNP Markers for Panicle Architecture and Grain Traits Based on RDP1 Gwa-QTL and Available for japonica Rice Improvement

Georgia C Eizenga, Aaron K. Jackson, Melissa H. Jia, Jeremy D. Edwards and Anna M. McClung, USDA-ARS Dale Bumpers National Rice Research Center, Stuttgart, AR

Panicle architecture, grain size and grain weight are important yield component traits to consider when breeding rice (*Oryza sativa* L.). Having SNP markers for these traits would expedite breeding efforts through marker assisted selection. Previously, the Rice Diversity Panel 1 (RDP1), representing the five major rice subpopulations, was phenotyped for these yield related traits and genome-wide association mapping (GWA)-QTL were identified. Most southern U.S. cultivars are classified as *tropical japonica* and California cultivars as *temperate japonica*; thus, these subpopulations are of particular interest to U.S. breeders. To develop markers and dissect the variation underlying these GWA-QTLs for rice improvement, diverse *japonica* RDP1 accessions were selected as parents to develop

biparental recombinant inbred line (RIL) mapping populations. All 276 progeny from the Estrela (admixture of *japonica*) × NSFTV199 (*tropical japonica*) cross have been evaluated for panicle architecture and grain traits. The QTL analyses with 256 RILs revealed 38 RIL-QTL which overlapped with the previously identified GWA-QTL and regions on chromosomes 3, 4, 5, 6, 7, 8 and 9 were selected for marker development. To develop markers, the sequence variation in regions surrounding the significant SNPs identified in GWA studies was assessed in rice genomic databases to find optimum sites to target for marker development. Once developed, the markers were tested in the Estrela × NSFTV199 population for amplification and polymorphism, and in two additional biparental *japonica* populations developed from RDP1 accessions. Polymorphic marker data were analyzed in these populations to identify marker-trait associations. Six Estrela × NSFTV199 RILs with desirable panicle architecture traits along with acceptable grain size, maturity, plant height and grain yield were selected for evaluation in a replicated field trial conducted over two years. The targeted SNP markers corresponded well with the panicle architecture and grain size data from the field studies. These SNP genotypes will be used to select RILs with the best agronomic and grain quality traits that are suited for the U.S. market along with QTL for desirable panicle architecture traits. The selected RILs will be released as improved germplasm for U.S. breeders.

PO0997: Legumes, Soybean, Common Bean, and related

Genetic Variation Underlying Seed Oil Content in Soybean

Yan Li, National Center for Soybean Improvement, Nanjing Agricultural University, Nanjing, China

Soybean is recognized as the largest oilseed crop that contributes to 61% of the world oilseed production and 28% of the total global vegetable oil consumption in 2018. The seed oil content has been increased during soybean improvement from landraces to released cultivars, and some genetic loci controlling oil content are likely subjected to selection. However, the genes underlying the selection are largely unknown.

Here, a population of 382 cultivated soybean accessions, including 187 landraces and 195 released cultivars, was used to screen the loci associated with seed oil content and likely subjected to selection. We identified 47 putative improvement-selective SNPs, and 25 of them overlap with the previously reported QTL for seed oil content. The 864 genes within the LD decay distance of these 25 SNPs were examined for their expression levels in different soybean tissues using RNA-seq data, and five genes were highly expressed in soybean seeds. Only one gene, *Glyma.15g049200* (*GmSWEET39*), showed much higher expression level in the seeds of high-seed-oil soybean variety than the low-seed-oil variety. Further analyses showed that the relative expression level of *GmSWEET39* was significantly correlated with soybean seed oil content. The sequence polymorphism in the promoter and coding region of *GmSWEET39* was significantly associated with oil content. The allelic effects of *GmSWEET39* on total oil content were confirmed in transgenic Arabidopsis, transgenic soybean hairy roots and soybean recombinant inbred lines. The frequencies of its superior alleles increased from wild soybean to soybean landraces, and are much higher in released cultivars. These findings suggest that the sequence variation in *GmSWEET39* affects its relative expression and oil content in soybean seeds, and *GmSWEET39* has been selected to increase seed oil content during soybean domestication and improvement.

In addition to *GmSWEET39*, the 864 genes in the 25 loci subjected to selective sweeps and locate within the previously mapped seed-oil QTL regions were further analyzed. The genes which were predicted to be involved in lipid metabolism related pathways were identified, and their roles in fatty acid accumulation should be investigated in future researches. Furthermore, among the SNPs which showed significant association with seed oil content by GWAS, many of them have not been selected during soybean improvement. The superior alleles and their germplasm carriers identified in this study would be valuable resources for the genetic improvement of seed oil content in soybean breeding program.

PE0998: Legumes, Soybean, Common Bean, and related

African-Led Genome Sequencing of Lablab and African Yam Bean Orphan Crops Genomes to Unveil Pathways Underlying Key Traits

Bernice Waweru¹, Jean-Baka K. Domelevo Entfellner², Oluwasheyi Shorinola³, Peter Emmrich⁴, Chris Jones³, Isaac Njaci⁴, Beatus Lymo⁵, Davies Kaimenyi⁶, Mukani Moyo-Okoba⁷, Edwin Murungi⁸, Ermias

Assefa⁹, Ahadi Birindwa¹⁰, Mary Gathoni Maranga¹¹, Abubaker Muwonge¹², Piuz Mugaga¹³, Hellen Nigussie¹⁴, Hassan Ali¹⁵ and Henry Osamede¹⁶, (1)KALRO, Njoro, Kenya, (2)Biosciences eastern and central Africa, International Livestock Research Institute, Nairobi, Kenya, (3)ILRI, Nairobi, Kenya, (4)BecA-ILRI Hub, Kenya, (5)Nelson Mandela Institute of Science and Technology, Tanzania, United Republic of, (6)The Sainsbury Laboratory, United Kingdom, (7)International Potato Center, Kenya, (8)Egerton University, Kenya, (9)Ethiopian Institute for Agricultural Research, Ethiopia, (10)Evangelical University in Africa, Congo (The Democratic Republic of the), (11)Jagellonian University, Poland, (12)NaCCRI, Uganda, (13)Makerere University, Uganda, (14)Addis Ababa University, Ethiopia, (15)University of Nyala, Sudan, (16)University of Ibadan, Nigeria

Legumes form a crucial part of diets in many Sub-Saharan Africa countries due to their high seed protein content and their low cost when compared to animal-derived protein sources. Despite offering excellent opportunities for sustainable intensification of agriculture, many legumes, especially orphan legume crops have received little research and breeding attention, causing large yield gaps and wasted potential for addressing the challenges of food security and sustainability. Two of these orphan legumes are lablab (*Lablab purpureus*) and African yam bean (AYB; *Sphenostylis stenocarpa*). Rich in protein, macro and micro-nutrients, both legumes are important for their edible leaves and seed grains. The African Yam bean also uniquely produces edible tubers while the haulms of both are utilized as animal feed. Both legumes exhibit nitrogen-fixing ability and thrive in marginal soils under low-input farming systems. However, both legumes are still largely underutilized, mainly due to their hard seed coats and the presence of anti-nutritional factors reducing digestibility. Little genomic information is available to assist in efforts to understand the genetic architecture underlying important traits and unlock the full potential of these crops. This study reports the first draft genomes of the AYB and lablab based on third generation long reads using Oxford Nanopore sequencing. Several approaches were used to generate long read *de novo* assemblies from 5.1 million AYB reads and 31 million lablab reads. Assembly using Redbean yielded an assembly length of 841 Mb with N50 of 48,083 bp for AYB and 354 Mb with N50 of 562,331 bp for lablab. Flye gave an assembly of 653 Mb for AYB with an N50 of 409,006 bp and 369 Mb with N50 of 1,442,401 bp for lablab. We also performed hybrid assembly using Illumina short reads to improve the accuracy of the assemblies and this yielded an assembly length of 486 Mb with an N50 of 7,124 bp for AYB. Efforts are ongoing to improve the contiguity of these assemblies to achieve chromosome-scale assemblies using Hi-C mapping. We are also generating RNA-Seq data for functional annotation of the AYB genome to identify key pathways underlying important nutritional, adaptation and resource partitioning traits.

This is the first report of the AYB genome, and long read lablab genome sequenced, assembled, and analyzed entirely in Africa by a group of young bioinformaticians following an ambitious capacity building effort at the Biosciences East and Central Africa ILRI Hub in Kenya. Our work therefore highlights the opportunities presented by orphan crop genome sequencing efforts for capacity building in agricultural genomics and bioinformatics in Africa.

PO0999: Fruit Species

Genomic Characterization of Novel Fruit Ripening Pathways

Seanna Hewitt, Washington State University, Pullman, WA, Christopher Hendrickson, National University, San Diego, CA, Evan Stowe, Molecular Plant Sciences, Washington State University and Amit Dhingra, Molecular Plant Sciences, Washington State University, Pullman, WA

Climacteric fruits are characterized by a dramatic increase in autocatalytic ethylene production, which is accompanied by a spike in respiration, at the onset of ripening. The change in the mode of ethylene production from autoinhibitory to auto-stimulatory is known as the system 1 (S1) to system 2 (S2) transition. European pear (*Pyrus communis* L.) cultivars require a genetically pre-determined duration of cold-temperature exposure to induce autocatalytic system 2 ethylene biosynthesis and subsequent fruit ripening. What happens during the cold treatment is becoming clearer at the molecular level. Differential expression, functional annotation, and gene ontology enrichment analyses have provided interesting evidence for the involvement of cold-induced, vernalization-related genes and repressors of endodormancy release, and an unexpected involvement of AOX transcription at pre-climacteric stage. These genes have not previously been described to play a role in fruit during the ripening transition. Besides the need for cold, application of 1-methylcyclopropene in European pear irreversibly obstructs

the onset of system 2 ethylene production resulting in perpetually unripe fruit. 1-MCP is an ethylene receptor antagonist which blocks ethylene perception and downstream ripening responses. In pear, application of exogenous ethylene, carbon dioxide and treatment to high temperatures is not able to reverse the blockage in ripening. Activation of AOX via exposure of 1-MCP treated 'D'Anjou' pear fruit to glyoxylic acid has been shown to trigger an accelerated ripening response. Ripening is consistently evident in decrease of fruit firmness and onset of S1-S2 ethylene transition. Transcriptomic and functional enrichment analyses have helped in identifying genes and ontologies implicated in glyoxylic acid mediated ripening, including alternative oxidase, TCA cycle, fatty acid metabolism, amino acid metabolism, organic acid metabolism, and ethylene responsive pathways. These data point to the glyoxylate cycle as a metabolic hub linking multiple pathways to stimulate ripening through an alternate mechanism. The results have provided information regarding how blockage caused by 1-MCP may be circumvented at the metabolic level, thus opening avenues for consistent ripening in pear and possibly other fruit. Understanding metabolic intervention points at which ripening responses can be manipulated provide key, species- and cultivar-specific gene targets which can be altered via gene editing or transgenic approaches for proactive modulation of ripening to enable development of strategies or new cultivars for reducing overall postharvest wastage.

PE1000: Fruit Species

Coconut Genetics and Genomics: Updates and Opportunities Towards a Vibrant Coconut Industry

Hayde C. Flandez-Galvez¹, Darlon V. Lantican¹, Anand Noel C. Manohar¹, Consorcia E. Reano¹, Maria Luz J. Sison¹, Susan M. Rivera², Ramon L. Rivera², Roanne R. Gardoce¹, Ernesto E. Emmanuel², Ma. Genaleen Q. Diaz¹, Cristeta A. Cueto², Barbara L. Caoili¹, Melvin P. Dancel¹, Don Emanuel M. Cardona¹, Romnick A. Latina¹, Edward Cedrick J. Fernandez¹, Cris Q. Cortaga¹, Jomari C. Domingo¹, Frenzee Kroeizha L. Pammit¹, Cecille Anne L. Osio¹, Michelle S. Guerrero¹, Jerice P. Monge¹, Cris Ann M. Lim¹, Ana Clarissa M. Ambagan¹, Rei Jedial O. Guevarra¹, Angelica Kate G. Gumpal¹, Erica Joy Amarilla¹, Dominic P. Agudera¹, John Carlo O. Campomanes¹, Maria Salve C. Vasquez¹, Hidelisa P. Hernandez¹, Susan R. Strickler³ and Lukas A. Mueller³, (1)University of the Philippines Los Baños, Philippines, (2)Philippine Coconut Authority (PCA), Philippines, (3)Boyce Thompson Institute, Ithaca, NY

Philippines is the second world supplier of coconut by-products. The region has been threatened with devastating production constraints ranging from agro-climatic and weather calamities to widespread prevalence of disease/insect pests outbreaks, and increasing existence of non-bearing and senile palms in coconut plantations. To facilitate the development of resilient and outstanding varieties especially for added high-value traits, advancements in genomics and related technologies are harnessed towards their effective integration in a coconut breeding program. Coconut whole genome sequence reads were generated using 'Catigan Green Dwarf' (CATD) as the reference variety and combinations of advanced next generation sequencing (NGS) platforms. High quality genome assembly was generated and used to characterize adaptation and economically important genes i.e. candidate resistance genes, drought tolerance, productivity, and coconut oil related genes. Genome-wide and gene specific DNA markers are generated. A user-friendly database is being developed to house the coconut genome sequence data, gene/trait models and associated DNA markers.

Updates from the Philippines coconut genomics project will be presented. These include gene mining for host resistance against coconut scale insect (CSI) and screening for CSI least damaged coconut varieties, as well as characterization of coconut genes related to fruit flesh/endosperm mutations and coconut oil qualitative/quantitative traits. Significant result from initial molecular and biochemical studies that support nutritional and medicinal claims will also be presented. The unprecedented opportunities beyond basic science from these major S&T achievements in coconut and in integration with applicable new breeding technologies will be discussed.

PO1001: Microbes and Pathogens

Citrusgreening.Org: An Open Access and Integrated Systems Biology Portal for the Huanglongbing (HLB) Disease Complex

Mirella Flores-Gonzalez¹, Prashant S Hosmani¹, Noe Fernandez-Pozo², Marina Mann³, Jodi L. Humann⁴, Dorrie Main⁵, Michelle Heck⁶, Susan J. Brown⁷, Lukas Mueller⁸ and **Surya Saha**¹, (1)Boyce Thompson Institute, Ithaca, NY, (2)University of Marburg, Germany, (3)Cornell University, Ithaca, NY,

(4)Washington State University, Pullman, WA, (5)Department of Horticulture, Washington State University, Pullman, WA, (6)USDA ARS, Ithaca, NY, (7)Division of Biology - Kansas State University, Manhattan, KS, (8)BTI, Ithaca, NY

We have created an open access web portal with pathosystem-wide resources and bioinformatics tools for the host citrus, the vector Asian citrus psyllid (ACP) and multiple pathogens including *Ca. Liberibacter asiaticus*. To the best of our knowledge, this is the first example of a database to use the pathosystem as a holistic framework to understand an insect transmitted plant disease. This endeavor integrates and enables the analysis of data sets generated by the community to study the citrus greening disease complex. Users can submit relevant data sets to enable sharing and allow the community to better analyze their data within an integrated system. The portal contains a variety of tools for omics data. Metabolic pathway databases, DiaphorinaCyc and CitrusCyc provide organism specific pathways and can be used to analyze transcriptomics and proteomics results to identify pathways with differentially regulated genes. Psyllid Expression Network (PEN) contains expression profiles of ACP genes from multiple life stages, tissues, conditions and hosts. Citrus Expression Network (CEN) contains public expression data from multiple tissues and conditions for citrus from NCBI. All tools like Apollo/JBrowse, Biocyc, Blast, CEN and PEN connect to a central database containing gene models for citrus, ACP and multiple *Liberibacter* pathogens. The portal also includes electrical penetration graph (EPG) recordings of ACP feeding on citrus and metabolomics data in addition to traditional omics data types with a goal of combining and mining all information related to a pathosystem. The portal includes user-friendly manual curation tools to allow the research community to continuously improve this knowledge-base as more experimental research is published. Bulk downloads are available for all genome and annotation datasets from the FTP site (<ftp://ftp.citrusgreening.org>). The portal can be accessed at <https://citrusgreening.org/>. More information can be found in the preprint at <https://www.biorxiv.org/content/10.1101/868364v1>

PE1002: Microbes and Pathogens

Variability and Plasticity of the Wild Chickpea Microbiome

Betsy Alford and Douglas R Cook, University of California-Davis, Davis, CA

Microbiomes of domesticated crops have come under increasing study for their potential to provide services to plants such as protection from biotic and abiotic stressors and in turn stabilize yields. While significant effort has been made to improve crop microbiome health, little is known regarding the “ground state” of the plant population microbiome. Here, we examine the microbial community structure of the wild progenitor of chickpea, *Cicer reticulatum*, and a sister species, *C. echinospermum*, across its projected native range.

As part of this work, we surveyed 20 sites throughout southeastern Turkey for soil characteristics, plant genetics, and microbial community assortment. In these locations of long standing co evolution, we observe consistent taxonomic microbial guilds corresponding with soil chemistry and plant species. To disentangle these relationships, sites with distinct soil chemistries and plant populations were chosen for a reciprocal transplant experiment. While the wild chickpea roots consistently enriched for microbial guilds regardless of soil type, the exact taxonomic membership is correlated with plant population. Conversely, nodule guild membership does not exhibit a clear plant population signal demonstrating that this relationship may be driven by the availability of compatible microbes in the soil.

PO1003: Methods: Bioinformatics

Eukaryotic Gene Prediction By Genemark-EP+ with Support from Homologous Proteins

Alexandre Lomsadze, Tomas Bruna and Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA

One of long-standing and difficult to reach goals of computational genomics is development of an accurate and fast algorithm for gene prediction in eukaryotic genomes. We present a new automatic gene finding tool, GeneMark-EP+, that integrates footprints (hints) of spliced aligned to genome homologous proteins into iterative unsupervised process of model training and gene prediction. A novel specialized pipeline, ProHint, generates high specificity hints for positions of intron borders as well as translation starts and stops by mapping homologous proteins from multiple species. For example, in *Arabidopsis thaliana*, with proteins of species outside *A. thaliana* taxonomical

order, ProtHint retrieves 70% of all introns with 99% specificity. GeneMark-EP+ uses the hints to improve estimation of model parameters as well as to directly adjust gene prediction with guidance of the most reliable hints. We tested GeneMark-EP+ on genomes of both model and non-model organisms. The results demonstrated improvements in prediction of exon-intron structures, particularly in large eukaryotic genomes, even when only proteins from remote species (outside the species phylum) are available.

PE1004: Methods: Bioinformatics

Improving Methods of Automatic Annotation of Plant Genomes

Tomas Bruna, Alexandre Lomsadze and **Mark Borodovsky**, Georgia Institute of Technology, Atlanta, GA

Integration of different pieces of evidence, i.e. RNA-Seq reads and homologous protein mapping and *ab initio* derived sequence patterns is critical for accurate annotation of plant genomes. Complex plant genomes have large size, large number of repeats (transposable elements) as well as heterogeneous nucleotide composition.

It was shown that when general gene prediction methods were applied to plant genomes significant manual work is still needed to reach genome annotation with satisfactory accuracy. Construction of a fully automated method of annotation of novel complex genomes is still an open problem.

Earlier we have developed automated gene finding method with unsupervised training of statistical models employed in the algorithm. This type of approach generates many false positives in complex plant genomes with large volume of non-coding regions. We present a novel method of model training on which sets of coding and non-coding regions is selected based on mapping of transcriptome and protein data.

New algorithm, GeneMark-ETP+, integrates RNA-Seq short read alignments produced by VARUS as well as hints generated by protein mapping delivered by ProtHint. This semi-supervised training approach was shown to generate more accurate annotation of principle isoforms of protein coding genes. A focus of the new method is selection of a highly reliable set of introns derived from RNA-Seq reads, proteins, and *ab initio* predictions to guide *ab initio* gene finder training and predictions. We demonstrated that this approach resulted in increase in accuracy of gene annotation in complex plant genomes.

PO1005: Methods: Bioinformatics

Analyzing Population Signatures of Cattle Antibody Repertoires

Yana Safonova¹, Luke Kramer², Sung Bong Sing³, James M. Reecy², Pavel Pevzner⁴ and Timothy P.L. Smith⁵, (1)University of California San Diego, La Jolla, CA, (2)Department of Animal Science, Iowa State University, Ames, IA, (3)USDA, (4)Computer Science Dept., UCSD, La Jolla, CA, (5)USDA, ARS, USMARC, Clay Center, NE

Repertoire sequencing (or *Rep-Seq*) technologies enable high-throughput scanning of antibody repertoires and open up new horizons for analyzing properties of adaptive immune system with downstream applications in the developing antibody drugs, estimating vaccine efficacy, and analyzing genomic diversity of immunoglobulin loci. Rep-seq made possible sequencing antibody repertoires of non-human species, including agriculturally important animals like cows. Immunoglobulin heavy chain locus of the cattle genome contains unusually long diversity (D) genes (~150 nt in cows vs ~30 nt in humans) that contribute to production of ultra-long CDR3s. Although recent studies have demonstrated the biomedical potential of antibodies with ultra-long CDR3s in treatment of HIV infection, their role in cattle antibody response remains unclear. In this work, we analyze cattle antibody repertoires taken before and after vaccination against BRDC (Bovine Respiratory Disease Complex) from 200+ individuals. We inferred alleles of variable (V) genes using Rep-seq data and found that genomic variations in V genes are strongly associated with the fraction of ultra-long CDR3s and the antibody production before and after the vaccination.

PE1006: Cotton

Development of Two Sets of CSSLs with *G.hirsutum* × *G.Barbadesnse* and Identification of Genetic Effects Using Slaf-Seq

Youlu Yuan, Institute of cotton research of CAAS, ANYANG, China

Development of Two Sets of CSSLs with *G.hirsutum* × *G.barbadesnse* and Identification of Genetic Effects using SLAF-seq

Shao-qi Li¹, Ai-ying Liu¹, Ju-wu Gong¹, Jun-wen Li¹, Wan-kui Gong¹, Qun Ge¹, Quan-wei Lu², Peng-tao Li², Yu-zhen Shi^{1*}, You-lu Yuan^{1*}

1 State Key Laboratory of Cotton Biology, Key Laboratory of Biological and Genetic Breeding of Cotton, The Ministry of Agriculture, Institute of Cotton Research, Chinese Academy of Agricultural Science, Anyang, 455000, Henan, China;

2 School of Biotechnology and Food Engineering, Anyang Institute of Technology, Anyang 455000, Henan, China.

*Correspondence: yuanyoulu@caas.cn (YYL) or shiyzmb@126.com (YZS)

Abstract

Detecting and mapping important chromosome regions related to quantitative phenotypic variation in Chromosome Segment Substitution Lines (CSSLs) in *Gossypium hirsutum* (*G.h*) background from *Gossypium barbadense* (*G.b*) provides an effective means to characterize the genetic basis of complex agronomic trait. Here, we developed two sets of CSSLs consisting of 408 and 332 lines, which were derived from Hail(*G.b*) × CCRI36(*G.h*) (BC₅F_{3:5}) and Hail(*G.b*) × CCRI45(*G.h*) (BC₄F_{3:5}), respectively. All of the two sets of CSSLs were subjected to high-throughput genotyping by Specific-Locus Amplified Fragment Sequencing (SLAF-seq) to obtain accurate physical maps based on the reference genome of *G. hirsutum* acc. TM-1_V2.1 (ZJU). We selected the 364--444bp genomic DNA fragments digested by *Rsa*I and *Hae*III as SLAF tags. A total of 440,081 SLAF tags were predicted to be uniformly distributed on the genome. And 4,316M reads (926.66 Gb) and 3,997M reads (868.14 Gb) were obtained from the two populations, respectively. After genotyping, 31,721 of 19,984,651 and 27,757 of 19,735,332 high quality SNPs were screened, with the average sequencing depths of 62.01X and 42.31X, respectively. And 13,974 SNPs were shared by the two sets of CSSLs. The genome was divided into 5,825 bins with the average length of 394.85 kb and 5,022 bins with the average length of 457.98 kb, respectively. The two sets of CSSLs were used to analyze variation for 9 agronomic traits in multiple environments and detect 1371 QTL linked to 691 bins and 1807 QTL linked to 691 bins, respectively. A total of 140 and 165 QTL were stably detected under 3-10 environments for the same traits, respectively. Among the bins with QTL, 429 (62.08%) for CCRI36 CSSLs and 339(41.01%) for CCRI45 CSSLs had overlapping intervals, which constituted of 212 common effect fragments. The results indicated that these chromosome fragments with common QTL in different genetic background may be the major genetic effect fragments from *G.b* to *G.h*, while the chromosome fragments with stable QTL in multiple environments in specific genetic background may interact with the genetic background.

Keywords: CSSLs, Chromosome introgressed segments, Genetic effects, SLAF-seq

PO1007: Fruit Species

Degs Involved in Formation of Pecan (*Carya illinoensis*) Pistillate and Staminate Flowers

Hormat Shadgou¹, Yiyi Li², Sajal Kumar², Joe Song¹, Richard Heerema¹, Jeremy Schmutz³, Jerry Jenkins³, Jane Grimwood³, L.J. Grauke⁴ and Jennifer J. Randall¹, (1)New Mexico State University, Las Cruces, NM, (2)New Mexico State University, (3)HudsonAlpha Institute for Biotechnology, Huntsville, AL, (4)USDA ARS Pecan Breeding & Genetics, Somerville, TX

Pecan [*Carya illinoensis* (Wangenh.) K. Koch] is a monoecious tree in the Juglandaceae family and is indigenous to North America. Pecan is one of the most economically important nut-producing species in the U.S. Alternate bearing (AB), the year-to-year fluctuations in nut production, is one of the horticultural restraints for the pecan

industry. AB in pecan is largely driven by fluctuations in the production of pistillate flowers. To address the genetic mechanism of alternate bearing in pecan, more than 52 RNA libraries including different varieties and tissues from several time points were collected and sequenced. RNA libraries were made from bud, pistillate flower and catkin tissues. Initial analyses of the samples from protandrous and protogynous trees at different time points through the growing season revealed a few candidate genes that are possibly involved in the timing of staminate flower formation. This data shows significant differences in the number of differentially expressed genes among all the samples specifically between the samples from the two different types of varieties collected in June and March. Differential gene expression of known flowering genes is also significant between the two different pecan varieties in different time points revealing the genes involved in pistillate and catkin formation. This data potentially reveals the gene networks involved for the formation and development of pecan pistillate flowers which could mitigate alternate bearing.

PE1008: Legumes, Soybean, Common Bean, and related

Global Gene Coexpression Networks Give Insight into the Evolution of Nodulation in Non-Legumes and Legumes

Xuelu Wang, Huazhong Agricultural University, Wuhan, China

Legumes can form nodules by intimate symbioses with rhizobia to fix atmospheric nitrogen. Symbiotic nitrogen fixation (SNF) provides about 40-60 million tons of nitrogen for agricultural systems each year in an ecofriendly manner. What happens for SNF during the evolution of legumes is always an outstanding question in evolutionary developmental biology. Soybean (*Glycine max*) is the most important crops for proteins and dietary oil and the fourth largest crop in production in the world. In addition, soybean belongs to ureide-forming legume and has a high efficiency of translocation of fixed N as ureide. We used weighted coexpression network analysis (WGCNA) to mine a nodule-related module (NRM) in soybean. Comparative genomic analysis of 78 green plant species revealed that NRM genes are recruited from different evolutionary nodes along with gene duplication events. A set of core coexpressed genes within legumes may play vital roles in regulating nodule environments essential for nitrogen fixation, including oxygen concentrations, sulfur transport, and iron homeostasis (such as *GmCHY*). We revealed that ancient orthologs and duplication events before the origin of legumes were preadapted for symbiosis. Conserved coregulated genes found within legumes paved the way for nodule formation and nitrogen fixation. These findings provide significant insights into the evolution of nodulation and indicate promising candidates for identifying other key components of legume nodulation and nitrogen fixation. We are currently working on the hub genes in the NRM to investigate their role in soybean nodule development.

PO1009: Legumes, Soybean, Common Bean, and related

Genome-Wide Association Analysis of Resistance to Scab (*Sphaceloma* spp) in Cowpea (*Vigna unguiculata* (L) Walp) in Uganda.

Abel Elisante Ngowi¹, Paul T Gibson², Patrick Obia Ongom³, Richard Edema², Joyline Rukarwa Runyararo⁴ and **Isaac Onziga Dramadri²**, (1)Department of Agricultural Production, College of Agricultural and Environmental Sciences (CAES), Makerere University, Uganda, (2)Makerere University Regional Centre for Crop Improvement (MaRCCI), Kampala, Uganda, (3)IITA (International Institute of Tropical Agriculture), Kano, Nigeria, (4)The Regional Universities Forum for Capacity Building in Agriculture (RUFORUM), Uganda

Cowpea (*Vigna unguiculata* L. Walp) is an important grain legume crop consumed in most parts of sub Saharan Africa. However, cowpea scab (*sphaceloma* Sp.) is one of the fungal diseases that causes significant yield loss under farmer condition. Genome-wide association study (GWAS) was conducted to determine the genetic architecture of scab resistance in 200 genotypes of cowpea from the mini core collection. The panel was evaluated under field conditions at two locations for two seasons in Uganda and genotyped with 50K single nucleotide polymorphism (SNP) markers. Nine significant association signals were identified for scab resistance under field conditions. The identified genomic regions and candidate genes once validated, can be used for initiating marker-assisted selection for developing cowpea varieties with resistance to scab disease in Uganda. The study provides insights into the genetic control to scab resistance in Cowpea.

PE1010: Legumes, Soybean, Common Bean, and related

Transcriptome Profile Reveals Drought Induced Genes Preferentially Expressed in Response to Water Deficit in Peanut (*Arachis hypogaea* L.) Genotypes

Xu Wang¹, Xinlei Yang², Phat M. Dang³, Rita M. Graze¹, Yucheng Feng¹ and **Charles Y. Chen⁴**, (1)Auburn University, (2)Agricultural University of Hebei, China, China, (3)USDA-ARS National Peanut Research Lab, (4)Department of Crop, Soil and Environmental Sciences, Auburn University, Auburn, AL

Cultivated peanut (*Arachis hypogaea*) is one of the most widely grown food legumes in the world being valued for its high protein and unsaturated oil contents. Drought stress is one of the major constraints that limits peanut production. The objective of this study was to identify the drought responsive genes preferentially expressed under drought stress in different peanut genotypes. To accomplish this, four genotypes (drought tolerant: 'C76-16' and '587'; drought susceptible: 'Tifrunner' and '506') used in a rainout shelter experiment were examined. Whole-transcriptome sequencing analysis identified 7,780 genes differentially expressed in Tifrunner and 9,767 in '506'. Of the 7,780 genes in Tifrunner, 5,310 genes were up-regulated and 2,470 were down-regulated. For the drought tolerant genotypes, 12,348 DEGs were identified in '587', including 7,172 up-regulated genes and 5,176 down-regulated genes. In C76-16, a total of 13,005 DEGs were identified with 7,718 up-regulated genes and 5,287 down-regulated genes. A total of 2,457 DEGs were shared by all four genotypes. Functional analysis of the shared DEGs identified a total of 139 enriched gene ontology (GO) terms consisting of 86 biological processes, and 53 molecular function, and defense response, reproductive process and signaling pathways were significantly enriched. Total of 43 significantly enriched Kyoto encyclopedia of genes and genomes (KEGG) pathways were also identified, and the most enriched pathways are those process involved in metabolic pathways, biosynthesis of secondary metabolites, plant circadian rhythm, phenylpropanoid biosynthesis, starch and sucrose metabolism, etc. This research expands our current understanding of the mechanisms that facilitate peanut drought tolerance and shed light on breeding advanced peanut lines to combat drought stress.

PO1011: Legumes, Soybean, Common Bean, and related

Marker Assisted Red Clover (*Trifolium pratense*) Cultivar Development for the Southern Great Plains

Kimberly Cooper, Erin Hubbard, Dusty Pittman, Wes Scruggs, Willy Inselman, Sarah A. Krogman and Michael Trammell, Noble Research Institute, LLC

Utilized in inter-seeding of hay pastures, grazing, and crop rotations, red clover is an important forage crop in the southern USA. The nitrogen fixing capabilities of the short-live perennial legume make it beneficial for increased soil health. This in addition to cheap and easy crop establishment along with high protein forage content make red clover a viable producer option for augmenting a multitude of cropping systems. However, in the Southern Great Plains, the inconsistency of red clover stand persistence can have a negative economic impact. Disease, environmental extremes, grazing tolerance and insect pressure can all affect stand persistence. Consequently, in 2016, the Noble Research Institute, LLC began a red clover evaluation of 30 GRIN (Germplasm Resource Information Network) accessions representing 15 different countries along with 6 commercial checks and 20 Noble developed synthetic populations to address traits linked to stand persistent in the southern Great Plains. Phenotypic data was taken, and grazing pressure applied on accessions over 3 consecutive years in multiple locations. Leaf samples, representing a pool of genotypes, were collected from each accession. Genomic DNA will be isolated from these leaf samples with the purpose of generating single nucleotide polymorphism (SNP) markers from a GBS (Genotyping-by-sequencing) method. A Genome Wide Association Study (GWAS), exploring marker-trait associations will be conducted. Selections from desirable accessions will be made based on both phenotypic and marker assisted selection (MAS) data to identify breeding genotypes for red clover cultivar development for the southern Great Plains.

PE1012: Legumes, Soybean, Common Bean, and related

Characterizing Resistance to Soybean Cyst Nematode in PI 494182, an Early-Maturing Soybean Accession

Vincent-Thomas Boucher St-Amour, Laval University, Quebec, QC, Canada

The soybean cyst nematode (SCN) generates more damage to soybean than any other parasite in most soybean-producing countries. The use of SCN-resistant cultivars remains the most effective method to limit losses caused by SCN. The SCN-resistant accession PI 88788 has been used almost exclusively to control SCN over the past decades, inducing a shift in nematode virulence to overcome the resistance. Furthermore, PI 88788 and other sources of resistance characterized to date belong to maturity groups (MGs) III and higher, making them less attractive to develop early-maturing soybean varieties (MGs 0-000). In this work, we performed a quantitative trait locus (QTL) analysis of the SCN-resistant soybean accession PI 494182 (MG 0). A recombinant inbred lines (RILs) population (Costaud x PI 494182) segregating for SCN resistance was challenged with SCN (Hg type 0) and genotyped via genotyping by sequencing (GBS) to produce a genetic map. Six resistance QTLs were identified, including a potentially new resistance locus on chromosome 07. A subset of the RIL population was confronted to a Hg type 2.5.7 SCN population and some of these exhibited resistance towards this type. Whole-genome sequencing of PI 494182 and Costaud allowed us to determine the alleles and their copy number for three candidate genes: *GmSNAP11*, *GmSNAP18* (*Rhgl*) and *GmSHMT08* (*Rhg4*). Finally, we determined that selecting for PI 494182 alleles at some SCN-resistance QTLs could entail linkage drag (decrease in protein content and 100-seed weight, increase in oil content). This work provides useful markers for introgressing SCN resistance in early-maturing soybean varieties.

PO1013: Legumes, Soybean, Common Bean, and related

Developing New Methods to Measure Traits Impacting Soybean Shoot Architecture

Suma Sreekanta¹, Riley C W O'Neill², Austin A. Dobbels¹, Jacob Meierhofer³, Fumiaki Katagiri⁴, Chehrzad Shakiban², Jeff Calder⁵, Peter Olver⁵, Aaron Lorenz¹, Robert M. Stupar¹ and Gary J. Muehlbauer¹, (1)Department of Agronomy and Plant Genetics, University of Minnesota, St. Paul, MN, (2)University of St Thomas, St Paul, MN, (3)College of Biological Sciences, University of Minnesota, St. Paul, MN, (4)Department of Plant and Microbial Biology, University of Minnesota, St. Paul, MN, (5)School of Mathematics, University of Minnesota, Minneapolis, MN

Shoot Architecture (SA) is a result of complex interplay between many traits. In crops such as maize and wheat, altering SA has resulted in enhanced yield. However, study of SA has been limited to a few traits because measuring SA traits has traditionally been a slow, low throughput process. Many previous studies have relied on manual measurements of a few traits that can be error prone as well as susceptible to measurement biases. We used a combination of high-throughput technologies including an unmanned aircraft system as well as inexpensive smartphone images to parameterize SA in terms of multiple individual leaf, branch and whole plant traits of field grown plants. A panel of 40 genotypes with strong visual variation in SA were used to evaluate our methods. Canopy coverage was measured each week using a drone based approach. SA traits were captured on images by both destructive and non-destructive sampling of plants. We have developed protocols on a MATLAB platform to automate image analysis. We automated processes to recognize and measure angle of branching, petiole angle, leaf shape characteristics, petiole length as well as overall plant length from images. We standardized tools and techniques to capture images that can be accessed by our image processing pipeline. Our automated image processing shows over 90% co-relation to manual measurements while reducing overall processing time significantly. Our methods for both image capture and image analysis rely on easily accessible equipment and are readily deployed by any lab interested in SA traits.

PE1014: Maize, Sorghum, Millet, Sugar Cane, and related

Annotation of the Tub Transcription Family Genes in Zea Mays

Sendi Meja¹, Marcela Karey Tello-Ruiz², Cristina F. Marco³, Doreen Ware², **Christos Noutsos**⁴ and Introduction to Bioinformatics class, (1)SUNY Old Westbury, (2)Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, (3)DNA Learning Center, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, (4)SUNY College at Old Westbury, Old Westbury, NY

This is a project that have been completed in a full academic year starting Spring semester 2019 and continuing to the Plant Biology in Fall semester 2019. The goal was to characterize the TUB Transcriptions Factors that are well conserved between animals and plants. In animals they play role in neuronal cells. In plants they might play role in

ABA signaling. In this work we have identified the 14 homologues of *Arabidopsis thaliana* in the *Zea mays* version 4 genome, using Gramene homolog search functions. In *Zea mays* the homologs are present in chromosomes 1 to 9. There is no TUB gene though in chromosome 10. This is a totally different organization compared with *Arabidopsis thaliana* where the majority of the 10 TUB genes are in chromosome 1. Another aspect of this work is the extend or alternative splicing we observed in Maize TUB seven out 14 TUB have alternative splicing events leading to proteins with no F-box or TUB domains. Based on RNA-Seq data the genes are expressed mainly in Roots. For that we have performed RT-PCRs to all of TUB Transcription Factors as well as in the alternative spliced products using cDNA from Roots and Shoots as well as Leaves to validate their structure.

PO1015: Methods: Bioinformatics

Panache : A Visualization Tool for the Exploration of Plant Pangenomes

Eloi Durant, IRD - UMR DIADE, France, Matthieu Conte, Syngenta Seeds, Toulouse, France, François Sabot, IRD - UMR DIADE, Montpellier, France and **Mathieu Rouard**, Bioversity International, Montpellier, France

High-throughput sequencing technologies enabled the production of multiple reference genome sequences for a single species. Comparisons of such sequences showed that there are structural variations between individuals from the same species such as Copy Number Variations (CNV) and Presence Absence Variations (PAV) that can have a significant impact on phenotypic variation in plants and could be suitable for breeding improved crop varieties. Thus, a single reference genome is insufficient to capture all variations.

Pangenomics is an integrative approach which aims to the assessment of such genomic variations and more within a group of closely related individuals. Its definition can focus on the whole repertoire of genes within a group or can include blocks of genomic sequences shared between species.

We introduce here a new visualization tool, based on a linear representation: the **PAN**genome Analyzer with **CH**romosomal **E**xploration (**PANACHE**). It is a web-based application which enables its users to explore a pangenomic reference divided in multiple panchromosomes. For now, it allows a quick identification of genomic blocks belonging to either the core or dispensable genomes with the representation of the corresponding Presence/Absence Matrix, and navigation within and between panchromosomes.

PE1016: Methods: Bioinformatics

Data Analysis and Visualisation Tools Supporting Users of DArT Service and Kddart Platform

Andrew Kowalczyk¹, Katarzyna Heller-Uszynska¹, Andrzej Kilian¹, Grzegorz Uszynski¹, Puthick Hok², David Hunter², Brad Lanham², Brian Pierce², Reuben Durham², Rick Spence² and Manil Chaudhary², (1)Diversity Arrays Technology Pty Ltd, Bruce, Australia, (2)Diversity Arrays Technology Pty Ltd, Australia

Diversity Arrays Technology (DArT PL) is an Australian organisation applying cutting edge genomics and Information Technologies to solve agricultural and environmental challenges. After a decade of focusing on DNA marker technology development and service, DArT has invested effort into developing a range of IT tools and expanding on a portfolio of analytical services and software. The KDDart platform (www.KDDart.org) is a suite of applications for storage, integration, collection and management of genotyping, phenotypic and environmental data. It services a range of small to medium size breeding companies and research teams involved in crop improvement and pre-breeding.

An integral part of KDDart is its analytical web application, named KDCompute, which offers a collection of user-friendly plug-ins to run downstream analysis using public-domain and DArT algorithms. Given our user base (breeders and geneticists) the main areas of development are around **Genome Wide Association Studies (GWAS)** and **Genomic Selection (GS)** as well as plugins for other analysis. KDCompute supports plain text file inputs or direct connection to a database via an API. One of the key plug-ins in the KDCompute framework is the DArTSoft (DS)14 application plugin, which performs automated genetic marker data extraction from sequencing libraries stored in DArTdb database.

To improve the quality of input data, KDDart suite offers a free Android app (KDSmart) for field data collection and KDXplore, an application for data curation and visualisation. We will present DArTview, a new marker data visualisation and filtering plugin in KDXplore, which enables more effective use of high volume marker data.

PO1017: Methods: Markers

Scalable and Cost-Effective Genotyping Technologies for Any Organism with Strong Analytical Support from DArT a Socially Minded Enterprise.

Kasia Heller-Uszynska, Jason Carling, Damian Jaccoud, Adriane Machado, Faisal Pirzad, Cina Vippin, Ling Xia, Puthick Hok, Grzegorz Uszynski and Andrzej Kilian, Diversity Arrays Technology Pty Ltd, Australia

Diversity Arrays Technology Pty Ltd since 2001 has been applying its cutting edge genomics and Information Technologies to solve agricultural and environmental challenges with its organic growth relying exclusively on “word of mouth” promotion. A culture of openness, consultancy and training underpins the provision of services to clients from over 60 countries, including many in the developing world.

Since 2010 DArT genotyping services have been based on New Generation Sequencing technologies. DArTseq™ is our highly scalable, whole genome analysis platform based on sequencing genomic representations generated using DArT genome complexity reduction methods. “Methyl filtration” embedded in the complexity reduction step ensures that both marker types (SNPs and SilicoDArTs) are within genic regions of the genome. The technology has been applied to over 1,000 species (bacteria, fungi, plants and animals) and its scalability based on library size and sequencing volume helps DArT to deliver genome profiles suited to particular application and budget.

Our Targeted Genotyping (TG) portfolio consists of DArTcap, DArTag and DArTmp assays, and delivers targeted SNPs based on markers derived either from DArTseq or from other platforms. These assays produce from 30 to over 10,000 targeted SNPs. The optimal platform and assay density/pricing are determined by user requirements.

Strong IT support is an integral part of the DArT genotyping services. With automated analytical pipelines covering each step from sequences to marker reports, and the KDDart platform for data management and downstream analytics (presented in the accompanying poster), DArT offers not just data, but rather solutions to problems.

PE1018: Tomato, Potato, Pepper, and related

Copy Number Variation in *Solanum bukasovii*

Ilayda Bozan¹, Maria Kyriakidou², Helen H. Tai³, Noelle Anglin⁴, David Ellis⁴ and Martina Stromvik⁵, (1)Department of Plant Science - McGill University, Canada, (2)Department of Plant Science - McGill University, Montreal, QC, Canada, (3)Agriculture and Agri-Food Canada, Fredericton, NB, Canada, (4)International Potato Center, Lima, Peru, (5)Department of Plant Science - McGill University, Ste. Anne de Bellevue, QC, Canada

Potato (*Solanum* sp.) is the world's most economically significant non-cereal crop. With a changing climate, and constant need for crop variety improvement, wild species are valuable genetic resources for breeding. While most cultivated potato varieties are tetraploid, the ploidy level varies among different potato species from diploid to hexaploid. Most of the tuber-bearing wild species are diploids. In this project, the genome of two wild diploid accessions of *S. bukasovii*, the closest wild species to the domesticated *S. tuberosum*, were studied. Whole genome sequence comparisons between *S. bukasovii* against other potato genomes, including two published reference genomes and a newly assembled diploid landrace was carried out to discover structural variation such as copy number variation between a wild species and cultivated potato.

PO1019: Tomato, Potato, Pepper, and related

Genomic Analysis of Heat Stress Tolerance during Tomato Pollination

Rob Reid, University of North Carolina Charlotte, Kannapolis, NC

Crop fertility loss due to extreme temperatures threatens the supply of food worldwide. Many factors contribute to heat sensitivity of flowering plant reproduction, including reduced partitioning of photosynthate to gametophyte development during chronic heat stress. However, even when vegetative and gametophyte development occur at optimal temperatures, acute heat stress that occurs in a brief window during the pollen tube growth phase has the potential to completely block seed production. Experiments on cotton provide one example, showing that temperature affects pollen tube growth in the pistil and that canopy temperature during pollination is a major determinant of crop production. Our goal is to focus on the pollen tube, pistil and the interactions between them to define networks of transcriptional, signaling, and physiological responses that underlie thermotolerance. This is a powerful system for fundamental exploration of how eukaryotic cells respond to stress because we have access to genome-scale tools that will focus on the pollen tube, a single cell with unique biology. Moreover, because we are using tomato pollen tube-pistil interactions as a model system, our work supports the urgent need to breed crop varieties adapted to elevated temperatures.

PE1020: Microbes and Pathogens

Pecan Seedling Microbiome Composition May be Influenced By Geographical Location and Host Genetics

Kimberly Cervantes¹, L.J. Grauke², Xinwang Wang², Patrick Conner³, Clive H. Bock⁴ and Jennifer J. Randall¹, (1)New Mexico State University, Las Cruces, NM, (2)USDA ARS Pecan Breeding & Genetics, Somerville, TX, (3)University of Georgia, Tifton, GA, (4)USDA-ARS Fruit and Tree Nut Research, Byron, GA

Native populations of pecan (*Carya illinoensis*) occur from Iowa, U.S.A. to Oaxaca, Mexico. Host genetics and environmental factors may play a role in microbiome composition, which can affect site adaptation and ultimately tree health. Therefore, it is important to elucidate the influence of geographical location and genetics on the microbial populations of pecan. To further investigate the bacterial and fungal composition of the pecan microbiome, we determined the microbial composition of seedlings from a controlled cross ('Lakota' × 87Mx3-2.11), grown from seed produced in two different geographical locations, Georgia and Texas. Seeds were planted in a soilless potting mix and grown in a quarantine facility. Total DNA was extracted from ten seedlings and sequenced using NGS technology. The resulting sequences were used to determine the bacterial and fungal compositions. Data was analyzed with the Qiagen CLC Microbial Genomics Module using the SILVA and UNITE databases. Initial 16S and 18S results indicate that all samples share many of the same classes of bacteria and fungi, respectively, but with some differences apparently based on geographical origin of the seed. For instance, the analyses revealed an association of *Methanomicrobia*, associated with brackish water such as found in GA, was identified in three of the GA samples while only in one of the TX samples. Recent innovations in microbiome analysis allow for the elucidation of the pecan microbiome composition which is important for understanding the contributions that the microbes make to the adaptation of this complex tree system.

PO1021: Microbes and Pathogens

The United States Swine Pathogen Database: Integrating Diagnostic Sequence Data of Emerging Pathogens of Swine

Blake T. Inderski, Tavis K. Anderson and Kay S. Faaberg, Virus and Prion Research Unit, National Animal Disease Center, USDA-ARS, Ames, IA

Objective: Veterinary diagnostic laboratories annually derive partial nucleotide sequences of thousands of swine pathogen isolates. In addition, next generation sequencing has resulted in the rapid production of full-length genomes. Presently, sequence data are released solely to diagnostic clients, as data are associated with sensitive information. However, this information can provide information to: objectively design field-relevant vaccines; determine when and how pathogens are spreading across the landscape; and identify virus transmission hotspots.

Methods: In tandem with the USDA-ARS Big Data initiative, we have developed a centralized sequence database at the National Animal Disease Center. We have implemented the Tripal v3 toolkit's BLAST interface and JBrowse genome visualization modules on Drupal v7 using the Chado database schema. Search forms supporting multi-

variable queries with customizable download options are available. Hosting is via Amazon Web Services for Federal Government with resource scaling and dedicated support.

Results: Sequences housed in the database contain at least four data fields: genomic information; date of collection; collection location (state level); and a unique identifier. Custom curation and annotation pipelines have been developed for multiple swine pathogens with capabilities of detecting the location of open reading frames, generating amino acid sequences, and identifying putative frame shifts.

Conclusion: The resource will provide researchers timely access to sequences discovered by veterinary diagnosticians, allowing for biological data mining and epidemiological studies. The result will be a better understanding concerning the emergence of novel viruses, how these novel isolates are disseminated in the US and abroad, and discovering new patterns of biological consequence.

PE1022: Fruit Species

Developmental Mutants of *Fragaria vesca*, the Diploid Woodland Strawberry

Janet P. Slovin, USDA/ARS Genetic Improvement of Fruit and Vegetables Lab, Beltsville, MD

Molecular research on the cultivated strawberry, *Fragaria x ananassa*, has focused on fruit development, primarily with respect to ripening and fruit quality, or resistance to disease, however, little is known about the molecular physiology of plant growth and development. Strawberry is a valuable perennial crop with a recently published genome sequence. Because *F. x ananassa* is octoploid, one of the diploid progenitors, the woodland strawberry *F. vesca*, has been developed as a model system for strawberry.

A high quality PacBio genome sequence of *F. vesca* is published, several inbred lines of this small plant are available, and *F. vesca* is transformable with *Agrobacterium*. Chemically induced mutants of *F. vesca* have been described, and CRISPR technology has been successfully applied this plant. To increase the utility of *F. vesca*, EMS treatment of imbibed seed was used to induce mutations in the runnering inbred line, Hawaii 4F7-3 (PI664444), the line used for sequencing. Resulting M2 seedlings and plants were scored for plant size and architecture, leaf color, texture and shape, flowering and flower morphology, ability to produce fruit, and fruit shape and size. Plants with mutations affecting runnering and crown architecture, as well as fruit shape have been further analyzed. Three mutants will be described here: 1) H4EMS703 is non-runnering, however the sequences of *FveGa20ox4* and *FveRGAI DELLA*, two genes reported as regulating runnering, are wild-type. Gene expression and bulked segregant analysis are underway with this mutant. 2) The fruit of H4EMS065 [*longfruit*, *lfi*] are long and slender. The *lfi* phenotype results from a single gene recessive mutation. Auxin application to the fruit throughout development has little effect on the *lfi* fruit phenotype. Analysis of *lfi* showed that fruit shape in strawberry is determined by the shape of the receptacle before the flower bud is open. 3) H4EMS068, has severely shortened inflorescences and also forms a large number of adventitious roots on runner plants.

PO1023: Maize, Sorghum, Millet, Sugar Cane, and related

Genomic, Metabolic, and Transcriptomic Responses of the Extremophile Grass *Paspalum Vaginatatum* to Nutrient Deficit Stress

Guangchao Sun^{1,2}, Nishikant Wase^{2,3}, Toshihiro Obata^{2,3} and James C Schnable^{1,2}, (1)Department of Agronomy and Horticulture, University of Nebraska-Lincoln, (2)Center for Plant Science Innovation, University of Nebraska-Lincoln, (3)Department of Biochemistry, University of Nebraska-Lincoln

Many wild grass species have been shaped by natural selection to thrive under environmental conditions or resource constraints far outside the range experienced by crop species. And improved understanding of the molecular and evolutionary strategies natural selection employed to achieve these changes in stress tolerance and nutrient use efficiency in crop wild relatives has the potential to aid efforts to engineer resilient and low input crops and advance food security. Here we focus on *Paspalum vaginatum* a crop wild relative which is multiple abiotic stress tolerant and exhibits greater tolerance to both nitrogen and phosphorous deficiency. *Paspalum* exhibits no significant decrease in biomass accumulation under nitrogen or phosphorous deficient conditions that significantly impact the biomass accumulation of maize and sorghum. All three species exhibit increased root elongation and branching in

response to nitrogen deficit. Metabolomic and transcriptomic analyses identified many commonalities in the molecular responses to stress in all three species. However, uniquely, paspalum exhibits significantly increases accumulation of trehalose under nutrient deficit conditions, and genes involved in metabolic pathways leading to trehalose production are experiencing more rapid protein sequence evolution in the lineage leading to paspalum than in other grass species. Efforts are underway to experimentally test the link between paspalum's unique strategy of accumulating trehalose in response to stress and paspalum's resilience to nutrient deficit.

PE1024: Wheat, Barley, Oat, and related

Combining Mutant Analysis and Genome Wide Association for Root Genetics Dissection in Barley

Silvio Salvi, Department of Agricultural and Food Sciences (DISTAL), University of Bologna, Bologna, Bologna, Italy

Knowledge about the genetic control of root development and architecture is lagging behind others plant traits and functions. This is negatively impacting modern plant breeding addressing drought tolerance and nutrient use efficiency. In this study, we investigated the genetic control of barley root architecture by exploring both induced and native genetic variation. We screened the barley TILLMore mutagenized population to identify root architecture mutants at seedling stages (2 weeks) using a semi-hydroponic system. We identified approx. 40 mutant lines, which grouped in three categories: root growth rate/length (short and long, 77%), root morphology (coiling or geotropic, 15%) and root hairs (hairless or shorthairs, 8%). Several mutants were tested for Mendelian inheritance and confirmed. SNP-based bulk-segregant analysis combined with exome and/or whole-genome shotgun sequencing enabled us to identify root candidate genes. Using the same root phenotyping protocol, a collection of >400 barley landraces and cultivars was phenotyped and GWA was carried out taking advantage of exome-seq based SNP analysis. Interestingly, the mutant loci and GWA-based QTL showed little if any overlaps, suggesting the presence of a largely undiscovered genetic system controlling root architecture in barley.

PO1025: Wheat, Barley, Oat, and related

Towards an Understanding of Beta-Glucan Regulation in Oat

Annis Fatmawati, Department of Plant Science, McGill University, Saint-Anne-de-Bellevue, QC, Canada

Oat is a major cereal crop that is grown worldwide for human food and animal feed. Its use in human diets is growing in popularity in part due to its ability to reduce serum cholesterol and glucose level. This has been attributed to its high β -glucan content. β -glucan is a major non-starch carbohydrate component, consisting of double β -1,3 and β -1,4 linkages. Information about β -glucan synthesis in oat is limited in part due to a lack of genomic resources. We are introducing the maize *Ac* and *Ds* transposable elements into the oat genome with the goal to create an experimental transposon-mediated functional genomic resource to identify genes encoding important traits such as β -glucan content. Recently, a Thaumatin Like Protein, TLP8 has been identified in barley that interacts with β -glucan to regulate its content in the grain. Higher transcript abundance of *TLP8* in barley grains reflect lower amounts of β -glucan, and vice-versa. We hypothesize that the downregulation of *TLP8* could increase β -glucan content in oat. The TLP8 homolog in oat was retrieved and an RNAi constructs created. Genetic transformation was then conducted via the bombardment gun method. Transformants were generated and selected using a phosphinothricin N-acetyltransferase (*PAT*) marker gene, yielding a 5-13% transformation efficiency. Histochemical assays confirmed the expression *PAT*, and transgenic plants were resistant to herbicide LIBERTY (0.2%). Currently, we are characterizing transgenic lines at the molecular and biochemical levels in order to explain the association between TLP8 and β -glucan in oat.

PE1026: Microbes and Pathogens

Sequence Capture NGS for Vertebrate Viruses

Oliver Lung, Canadian Food Inspection Agency, National Centre for Foreign Animal Disease, Winnipeg, MB, Canada

The availability of high through-put sequencing technology has improved our ability to detect and characterize pathogens without *a priori* knowledge. However, the abundances of host and environmental nucleic acids can impact the success of accurately identifying low abundant viral nucleic acids. ViroCap is a sequence capture panel which consists of probes designed to enrich sequences representing viral species that infect vertebrates which sequence information is available. ViroCap was designed to capture sequences from viruses representing 190 genera, and has been previously tested on 32 viruses representing 19 genera that affect humans. Here, we further tested ViroCap against a panel of blinded cell culture amplified viruses and clinical/field samples containing another 26 viral species representing 19 genera and 12 families that affects livestock, wildlife and humans. All viral species were accurately identified and a few unexpected viruses were detected. Full or near full genomes were obtained for most tested viral species and enrichment was observed when compared with pre-captured material. These results indicate ViroCap is a useful tool for improving the sensitivity of NGS for identification and sequencing of a broad spectrum of viruses that affect vertebrates.

PO1027: Rice

Development of Herbicide Tolerant Basmati Rice Lines through Marker Assisted Backcross Breeding

Nitasha Grover¹, Ashutosh K. Yadav¹, Gopala Krishnan S², Ranjith K. Ellur¹, Prolay K. Bhowmick¹, Viswanathan C¹, Nagarajan M³, Amitha Mithra S. V.⁴, R.P. Sharma⁵, Robin S⁶, Jeyaprakash P⁶, Raveendran M⁶, Sheshshayee M. Sreeman⁷, M. K. Kar⁸, N Sarla⁹, SK Mangrauthia⁹, N.K. Singh⁴, T Mohapatra¹⁰, **Ashok K. Singh¹** and Indian Agricultural Research Institute, New Delhi, India, (1)Indian Agricultural Research Institute, New Delhi, India, (2)Indian Agricultural Research Institute, New Delhi, India, New Delhi, India, (3)Rice Breeding and Genetics Research Centre, Aduthurai, Tamil Nadu, India, (4)ICAR-National Research Centre on Plant Biotechnology, New Delhi, India, (5)ICAR- National Research Centre on Plant Biotechnology, New Delhi, India, (6)Tamil Nadu Agricultural University, Coimbatore, India, (7)University of Agricultural Sciences, Bangalore, India, Bangalore, India, (8)National Rice Research Institute, Cuttack, India, (9)Indian Institute of Rice Research, Hyderabad, India, (10)Indian Council of Agricultural Research, New Delhi, India

Pusa Basmati 1121 (PB1121) is the landmark Basmati rice variety developed at ICAR-IARI, New Delhi, earns foreign exchange of US\$ 3.41 billion annually. The decrease in water availability coupled with labour scarcity has necessitated the shift in rice cultivation from transplanted rice to direct-seeded rice (DSR). Weed management is the major constraint in DSR which has a direct impact on productivity. Therefore, there is a need to develop herbicide tolerant (HT) rice varieties. We used marker assisted backcross breeding (MABB) to transfer the mutant allele of Acetohydroxy acid synthase (*AHAS*) gene, which confers tolerance to imidazolinonones group of herbicide from donor parent Robin into the genetic background of an elite Basmati rice variety, PB1121. Foreground selection was carried out using gene linked SSR marker RM6844 and background selection using 112 SSR markers polymorphic between PB1121 and Robin. Phenotypic selection for agronomic, grain and cooking quality traits was carried out in each of the generations to accelerate the recovery of recurrent parent phenotype. In BC₄F₄ generation, 12 near isogenic lines (NILs) with recurrent parent genome recovery ranging from 98.66 to 99.55% were isolated and evaluated under sprayed and unsprayed conditions. These NILs were either at par or superior to PB1121 under controlled conditions; and possessed no significant effect on yield, grain and cooking quality parameters under sprayed conditions. HT-PB1121 is expected to be the prime technology for rapid adoption of DSR in the Basmati growing areas.

PE1028: Other Plant Species

A Gene Expression Atlas of *Vellozia nivea*, a Desiccation-Tolerant Species from the Brazilian Campos Rupestres

Isabel R. Gerhardt, Embrapa/GCCRC, Campinas, Sao Paulo, Brazil, Vagner K. Okura, Lactad/Unicamp, Campinas, Sao Paulo, Brazil, Ricardo A. Dante, Embrapa/GCCRC, Campinas, Brazil and Paulo Arruda, Unicamp/GCCRC, Campinas, Brazil

Velloziaceae are an angiosperm family that contains the most desiccation-tolerant species (approximately 200 out of 270 species). More than 80% of the *Velloziaceae* species occur in South America, where the greatest morphological diversity is also found. The genus *Vellozia* comprises both desiccation-tolerant and non-desiccation-tolerant species, offering an excellent model for studying the evolution of desiccation- and drought-tolerance traits on plant genomes. To date, only limited genomic or transcript sequences are available for *Velloziaceae* species. Here we present a *Vellozia nivea* gene expression atlas across different plant organs and tissues, including flower, developing seeds, root, leaf, stem and seedling. *Vellozia nivea* is a desiccation-tolerant species, endemic to the Brazilian *campos rupestres* (rupestrian grasslands) and highly adapted to their extreme conditions. A total of 180.67 Gb of raw data were generated, and of these, 152.79 Gb were subjected to downstream analysis after quality control (QC). *Vellozia nivea* *de novo* transcriptome assembly was performed with the Trinity bioinformatics tool, resulting in 684.615 contigs. After filtering contaminated sequence contigs from bacteria and fungi and removal of contigs with less than 10 sequence reads associated with the initial assembly, the transcriptome resulted in 195.512 remaining sequences. A GO enrichment analysis was performed on tissue-specific transcripts. The *Vellozia nivea* transcriptome should be a useful resource for genome annotation and gene function discovery studies.

PO1029: Other Plant Species

Gingerroot: A Novel DNA Transposon Encoding Integrase-Related Transposase in Plants and Animals

Stefan H. Cerbin¹, Ching Man Wai², Robert Van Buren¹ and **Ning Jiang¹**, (1)Department of Horticulture, Michigan State University, East Lansing, MI, (2)Department of Horticulture, Michigan State University

Transposable elements represent the largest components of many eukaryotic genomes and different genomes harbor different combinations of elements. Here, we discovered a novel DNA transposon in the genome of the clubmoss *Selaginella lepidophylla*. Further searching for related sequences to the conserved DDE region uncovered the presence of this superfamily of elements in fish, coral, sea anemone, and other animal species. However, this element appears restricted to Bryophytes and Lycophytes in plants. This transposon, named *GingerRoot*, is associated with a 6 bp (base pair) target site duplication, and 100–150 bp terminal inverted repeats. Analysis of transposase sequences identified the DDE motif, a catalytic domain, which shows similarity to the integrase of *Gypsy*-like long terminal repeat retrotransposons, the most abundant component in plant genomes. A total of 77 intact and several hundred truncated copies of *GingerRoot* elements were identified in *S. lepidophylla*. Like *Gypsy* retrotransposons, *GingerRoots* show a lack of insertion preference near genes, which contrasts to the compact genome size of about 100Mb. Nevertheless, a considerable portion of *GingerRoot* elements was found to carry gene fragments, suggesting the capacity of duplicating gene sequences is unlikely attributed to the proximity to genes. Elements carrying gene fragments appear to be less methylated, more diverged, and more distal to genes than those without gene fragments, indicating they are preferentially retained in gene-poor regions. This study has identified a broadly dispersed, novel DNA transposon, and the first plant DNA transposon with an integrase-related transposase, suggesting the possibility of *de novo* formation of *Gypsy*-like elements in plants.

PE1030: Methods: Sequencing

"Seqoccin" : A Long Read Project to Find Optimal Technologic Combinations for Genome Assembly and Their Variability, Epigenetic Marks Detection and Metagenomic Analysis

Cécile Donnadiou¹, Camille Eché², Claire Kuchly³, Jérôme Lluch⁴, Olivier Bouchez¹, Carole Iampietro⁵, Christophe Klopp⁶, Adrien Castinel², Remy Serre², Celine Vandecasteele⁷, Amandine Broha², Andreea Dreau², Romain Therville², Gerald Salin², Didier Boichard⁸, Mekki Boussaha⁹, Alain C. Charcosset¹⁰, Frédérique Pitel¹¹, Julie Demars¹², Gwenola Tosser¹³, Clémentine Vitte¹⁴, Céline Lopez-Roques¹⁵ and Christine Gaspin⁶, (1)INRAE GeT Genomics Facility, Castanet-Tolosan, France, (2)GeT-PlaGe INRAE, France, (3)GeT-PlaGe Core Facility - INRA France, Castanet Tolosan cedex, France, (4)INRAE, Castanet, France, (5)GeT-PlaGe INRAE, Castanet-Tolosan, France, (6)MIAT, INRA, Université de Toulouse, Castanet-Tolosan, France, (7)GeT-PlaGe Core Facility - INRA France, Castanet tolosan, France, (8)INRA UMR1313 Animal Genetics and Integrative Biology, Jouy-en-Josas, France, (9)INRA UMR1313 Animal Genetics and Integrative Biology, Jouy en Josas, France, (10)INRA GQE - Le Moulon, Gif sur Yvette, France, (11)INRA, UMR 1388 GenPhySE, Castanet-Tolosan, France,

(12)INRA-GenPhySE-Génétique, Physiologie et Systèmes d'Elevage-UMR1388, Castanet-Tolosan cedex, France, (13)INRA GenPhySE, France, (14)GQE - Le Moulon, Gif sur Yvette, France, (15)INRA Plateforme Génomique GetPlaGe, Castanet-Tolosan, France

The SeqOccIn project (Sequencing Occitanie Innovation), supported by Get-PlaGe and Genotoul Bioinfo platforms, was selected by the Occitanie Region as part of the call for projects "Regional Research and Innovation Platforms". This project should enable us to acquire expertise on the optimal combination of long fragment sequencing technologies and associated applications to better characterize complex genomes in agronomical field: from SNP and structural variations detection, to the production of a high quality assembly at a lower cost. The analysis of native DNA molecules without amplification will allow the detection of some epigenetic marks, and the study of long fragments will allow us to go further on barcoding approaches, as well as on the sequencing of whole metagenomes. Our ambition is to combine and study the contribution of different technologies for three research axes: genome variability analysis, epigenetic mark analysis and metagenome.

PO1031: Maize, Sorghum, Millet, Sugar Cane, and related

Molecular Responses of Lowland and Upland Switchgrass Cultivars to Infection By Fungal Rust.

Nathan A. Palmer¹, Serge J. Edme¹, Gautam Sarath¹, Rob Mitchell², Satyanarayana Tatineni² and Gary Yuen³, (1)USDA-ARS Wheat, Sorghum, and Forage Research Unit, Lincoln, NE, (2)USDA-ARS Wheat, Sorghum, and Forage Research Unit, (3)University of Nebraska Lincoln

Sustainable, high biomass production from switchgrass (*Panicum virgatum*) is essential for its use as a biofuel feedstock. However, both biomass production and biomass quality are negatively impacted by pathogen attack. In particular, infection by *Puccinia spp* that cause fungal rust has been shown to significantly reduce ethanol yields of harvested switchgrass biomass. In general, lowland switchgrass cultivars tend to be more resistant to pathogens compared to upland switchgrass cultivars. A better understanding of the mechanisms behind this resistance will be advantageous for switchgrass breeders and enable the incorporation resistance traits into breeding populations. Towards this end, plants at the 4th leaf stage of lowland cultivar Kanlow and upland cultivar Summer were infected with spores of *Puccinia novopanici* (formerly referred to as *Puccinia emaculata*). Metabolic (metabolites and phytohormones measured by LCMS) and transcriptomic (RNA-Seq) responses relative to uninfected controls were analyzed across an 18-day time series in order to identify basal and infection response differences between Kanlow and Summer plants. Gene networks, enriched metabolic pathways, and key metabolites distinguishing the Kanlow and Summer responses to *P. novopanici* infection are presented.

PE1032: Fruit Species

Population Sequencing Reveals Clonal Diversity and Ancestral Inbreeding in the Grapevine Cultivar Chardonnay

Michael J. Roach¹, Daniel Johnson¹, Joerg Bohlmann², Hennie J. J. van Vuuren², Steven JM Jones³, Isak S. Pretorius⁴, Simon A. Schmidt¹ and Anthony R. Borneman¹, (1)Australian Wine Research Institute, Australia, (2)University of British Columbia, Canada, (3)BC Cancer Agency Genome Sciences Centre, Vancouver, BC, Canada, (4)Chancellery, Macquarie University, Australia

Chardonnay is the top white wine varietal worldwide and Australian Chardonnay's success on the world stage was underpinned by a historic program of clonal selection. But what is 'clonal selection'? Genetic mutations accumulate within a plant during successive propagations. This often causes phenotypic differences that alter yield, quality, or sensory characteristics. Favourable phenotypes can be captured and amplified by using a single plant with this phenotype as a master stock for new plantings. Clonal plantings can therefore improve vineyard performance or provide unique flavour and aroma profiles. Unfortunately, the genetics behind clonal differences is poorly understood.

We first produced a high-quality Chardonnay assembly using PacBio RS-II sequencing and FALCON Unzip. During this step we developed a tool--Purge Haplotigs--for reassigning the homologous primary contigs that remain after assembly of highly heterozygous genomes. This step is important for fully capturing heterozygous variants. We then sequenced 15 popular Chardonnay clones and identified 1,620 genetic markers that distinguish them, one of

which is a known Muscat mutation. Plants of the same clonal lines were then sequenced from independent source locations allowing the clonal markers to be validated. Clones were able to be reliably identified using these markers and regional differences were also identified within clonal populations.

Finally, we show that the Chardonnay genome contains extensive evidence of parental inbreeding, such that its parents, Pinot Noir and Gouais Blanc, may even represent first-degree relatives. This previously unreported finding sheds new light on the heritage of Chardonnay and Gouais Blanc.

PO1033: Forest Trees

Genetic Architecture of Adventitious Root and Related Shoot Traits in Populus By QTL Mapping and RNA Sequencing

Pei Sun, Huixia Jia and **Jianjun Hu**, Research Institute of Forestry, Chinese Academy of Forestry, Beijing, China

To dissect genetic architecture of adventitious root and related shoot traits and regulatory genes, we measured 12 adventitious root and related shoot traits of 434 F₁ population from *Populus deltoides* 'Danhong' × *Populus simonii* 'Tongliao1' by an integrative analysis of QTL mapping and RNA-Seq data. Extensive segregation, high repeatability, and significant correlation relationship were detected for the investigated traits. A total of 233 QTLs were associated with adventitious root and were associated with shoot traits, explaining 3.1–19.8% of phenotypic variation. Twenty-five QTL clusters and 40 QTL hotspots were identified for the investigated traits. Ten QTL clusters were overlapped in both adventitious root traits and related shoot traits. Transcriptome analysis identified 10,172 differentially expressed genes (DEGs) among two parents, three fine rooting and three poor-rooting genotypes, 143 of which were physically located within the QTL intervals. Combining the QTL and transcriptome analysis, three associated genes (*Potri.004G111400*, *Potri.T021600*, and *Potri.012G082800*) within the QTL intervals were differentially expressed between fine-rooted and poor-rooted genotypes. K-means cluster and weighted gene co-expression network analysis showed that *PtAAP19* (*Potri.004G111400*) encoding amino acid transport protein was tightly associated with adventitious roots and highly expressed in fine-rooting genotypes. Compare with 'Danhong', 153 bp deletion in the coding sequence of *PtAAP19* in 'Tongliao1' gave rise to lack one transmembrane domain, which might cause the variation of adventitious roots. Finally, this study deciphered the genetic basis of adventitious root and related shoot traits and provided potential function genes for genetic improvement of poplar breeding.

PE1034: Canine

Do Reproductive Barrier Genes Contribute to the Maintenance of Distinct Red Fox (*Vulpes vulpes*) Populations across a Zone of Secondary Contact?

Sophie Preckler-Quisquater, UC Davis - Mamm Ecol and Conserv Unit, Davis, CA and Benjamin N Sacks, UC Davis - Mammalian Ecology and Conservation Unit, Davis, CA

The Sacramento Valley red fox (SVRF, *Vulpes vulpes patwin*) is endemic to the northern Central Valley of California, where habitat loss has reduced its genetic effective population size and restricted its distribution. In the 1960s, a non-native population of red foxes was established to the south of the SVRF range from translocated individuals and escapees from fur farms. These two populations have been in contact for approximately 35-50 generations, yet remain genetically distinct except for a narrow hybrid zone that has remained stable over time. Given the apparent phenotypic similarities between these populations and the small extent of the SVRF range, which could be readily traversed by a single disperser, the long-term genetic distinctiveness of these two populations begs for a mechanistic explanation. One hypothesis is that through their respective evolutionary histories, that involve demographic bottlenecks and differential selective pressures, these populations have evolved Dobzhansky-Muller incompatibilities that reduce fitness of hybrid offspring. If so, we predict that relative to background genetic variation, genomic regions associated with sterility, many of which are on the X chromosome, would show greater differentiation between these populations. We used a genotyping-by-sequencing approach, sampling individuals from the native (n = 42) and nonnative (n = 52) ranges, and within the hybrid zone (n = 18). We identified >40,000 high-quality single nucleotide polymorphisms (SNPs), which we used to quantify levels of overall gene flow and to scan for statistically meaningful signals of differentiation associated with particular gene regions. We found highly

differentiated regions corresponding to sperm capacitation, embryo development, and olfaction. We also discovered a high density of differentiated regions on the X chromosome, including several involved in species-specific reproductive function. These findings support the hypothesis that reproductive barriers (e.g. reduced hybrid fitness) may contribute to the maintenance of these genetically distinct populations.

PO1035: Canine

Whole Genome Resequencing of Wolves at a Contact Zone in South Asia to Explore Distinctiveness, Local Adaptation, and Barrier Genes

Lauren Hennelly, Mammalian Ecology and Conservation Unit, UC Davis, Davis, CA

Widely distributed and vagile species can exhibit and sometimes maintain local adaptations despite high gene flow. One way this can happen is through the evolution of reproductive isolating mechanisms, which reduce gene flow near barrier genes. The gray wolf is distributed across Eurasia and North America, providing a good model to investigate these and other evolutionary processes. We utilized a contact zone between two independent and phylogenetically basal maternal wolf lineages corresponding to the Tibetan and Indian wolf, the latter of which had not been genomically characterized, along with Holarctic and Arabian populations. We sequenced the whole genomes from Tibetan and Indian wolves and combined these with previously sequenced Holarctic and Arabian wolves to evaluate (1) the overall genomic distinctiveness of Indian wolves, (2) potential genes under convergent selection in geographically distinct but environmentally similar Arabian and Indian wolves, (3) differing local selection between adjacent Tibetan and Indian wolves, which occupy distinct elevational and climatic environments, and (4) barrier genes limiting gene flow from the latter two taxa. We present preliminary results on the above 4 questions, highlighting portions of the genome previously associated with elevational and climatic adaptation and reproductive isolation. Our research will provide insight into the processes that influence population differentiation in highly mobile species and inform conservation priorities of these south Asian wolves.

PE1036: Brassicas, Arabidopsis, and related

Lead Tolerance and Accumulation in the Brassicaceae “*Hirschfeldia Incana*” : Advances in Molecular Mechanisms

Mouna Fahr¹, Florence Auguy², Laurent Laplace³, Mohammed El Mzibri⁴, Najib Bendaou¹, Valérie Hocher², Patrick Doumas², Didier Bogusz² and **Abdelaziz Smouni**⁵, (1)Université Mohammed V de Rabat, Rabat, Morocco, (2)Institut de Recherche pour le Développement, Montpellier, France, (3)Institut de Recherche pour le Développement, Montpellier, France, (4)Centre National de l’Energie, des Sciences et des Techniques Nucléaires, Rabat, Morocco, (5)Université Mohammed V de Rabat/ Institut de Recherche pour le Développement (IRD), Rabat, Morocco

Hirschfeldia incana, is a pseudometallophyte belonging to the Brassicaceae family, widespread in the Mediterranean region and a medium perennial shrub. This plant has good ability to grow on soils contaminated by lead (Pb). We used *Hirschfeldia incana* as a model for molecular characterization of Pb tolerance and accumulation in plants.

Microarrays comparison of gene expression between *H. incana* and a susceptible plant to Pb (*Arabidopsis thaliana*), enabled the identification of a set of specific genes expressed in response to lead exposure. The number of genes specifically regulated by Pb in *H. incana* was higher in shoots than in roots, with 602 and 341 genes, respectively. Four groups of genes were particularly over-represented due to the Pb exposure and are categorized into biological processes as photosynthesis, cell wall, transport and metal handling. Expression profile of a subset of differentially regulated genes involved specifically in transport and metal handling was also checked by qRT-PCR in *H. incana*.

Our results suggest a possible scenario of Pb tolerance and accumulation mechanisms in *H. incana*. In roots, Pb is probably chelated with defensins at the cytosol level and interact with membrane transporters. Then, HMA4 a P-type ATPase transporter may play a role in Pb translocation to aerial parts through the xylem. Once in the leaves, ferritin and Metallothionein can play an essential role in Pb distribution in leaves and Pb binding and sequestration in cells. MRP14 transporter can be involved in the Pb storage in the vacuoles probably in a conjugate form (glutathione-Pb). Another transporter, GCN2 from the ABCF family, can contributed in Pb storage in chloroplast.

This study provided new insights into the molecular mechanisms of Pb tolerance and accumulation in plants and opens new ways for the study and the selection of heavy metals tolerant plants.

PO1037: Legumes, Soybean, Common Bean, and related

Development of the First Molecular Linkage Map in Lima Bean: Synteny and Chromosomal Rearrangements Relative to the Common Bean Genome

Stephanie Zullo, University of California, Davis, Davis, CA and **Paul Gepts**, University of California, Davis, CA

Using genotyping by sequencing methods, a genetic map of lima bean was constructed with 3,752 SNP markers from a recombinant inbred line population (n=240) with a genetic distance of 1,142cM spanning a physical length of 506.7Mbp. Due to strong synteny between common bean and lima bean and the absence of a lima bean reference genome, alignment of the sequenced lima bean reads was made to a common bean reference genome (G19833, v2.1). Genetic divergence, chromosomal rearrangements, and recombination rate variation in the lima bean genome are being characterized relative to common bean using this genetic map. Distinct chromosomal rearrangements of the lima bean genome relative to common bean include confirmation of intra-chromosomal translocations on PI02 and PI09, pericentric inversions on PI09 and PI10, multiple inversions on PI03, PI04, PI07, duplication of a region of PI07 found on PI11, and chromosomal rearrangements and insertions on PI02, PI04, PI05, PI08 and PI11.

PE1038: Methods: Markers

Genome-Wide Association Studies Combining Genotyped and Non-Genotyped Relatives Using Bayesian Regression Methods with Mixture Priors

Abhilash Dhal, UC Davis, Davis, CA, India

Bayesian linear mixed models (BLMM) incorporating different mixture priors are widely used in genomic prediction. Such model architectures have been extended for combining genotyped and non-genotyped relatives in the pedigree in single-step Bayesian regression methods. In genomic prediction, these extended single-step Bayesian linear mixed models (SS-BLMM) yield similar or higher accuracy for genotyped individuals than could be obtained by ignoring phenotypic information from non-genotyped relatives. SS-BLMM can also be extended in the context of genome wide association studies (GWAS). Inference in GWAS using SS-BLMM can be based on posterior probabilities for genomic windows which capture joint signals of SNPs in a genomic region where a trait causal variant (i.e., QTL) exists. In this study, we present SS-BLMM by using window based posterior probability of association (WPPA) as a test statistic for GWAS inference. Our comparisons of SS-BLMM with BLMM across different genetic architectures (e.g., heritability, proportion of genotyped individuals in the population, number of QTL, and genomic region size) showed superior performance of SS-BLMM over BLMM. Heritability and proportion of animals genotyped in the population were found to have a significant influence on the performance of both models in the present study. SS-BLMM is able to achieve a significantly higher detection accuracy in comparison to BLMM for fixed window size analysis at lower heritability and lower proportion of animals genotyped. SS-BLMM can detect more true positive genomic windows as compared to BLMM across different heritabilities as well as different proportions of animals genotyped. In addition, we also discovered that adaptive window inference showed higher accuracy than fixed window based inference across different fixed window sizes. The software tool JWAS offers open-source implementation to perform these GWAS analyses.

PO1039: Natural Populations

Genomic Consequences of Inbreeding and Outbreeding in an Endangered Carnivore Population

Malin Hasselgren, Stockholm University, Stockholm, Sweden

Loss of genetic variation through genetic drift and inbreeding is a major threat to small and isolated populations. In a subpopulation of the Scandinavian arctic fox (*Vulpes lagopus*), inbreeding depression and genetic rescue has recently been documented through pedigree analyses linked to life history traits. Isolation for nine years at an extremely small population size led to a rapid increase in inbreeding, with lower survival and reproduction in inbred

individuals as a result. However, an event of natural immigration by three outbred males, of which two were brothers, resulted in genetic rescue through elevated fitness in immigrant F1 offspring.

By sequencing complete genomes of 23 Scandinavian arctic foxes born before and after the immigration event, we here look into the genomic consequences of inbreeding and outbreeding over multiple generations. We found that immigrant F1 offspring had 18 % higher genome-wide heterozygosity and 52 % lower genomic inbreeding compared to native individuals. Heterozygosity and inbreeding in immigrant F2 and F3 did not differ from native individuals or immigrant F1. We also found that foxes surviving their first year generally had higher heterozygosity and lower inbreeding. Furthermore, genomic inbreeding levels correlated with pedigree based inbreeding coefficients, however, the pedigree consistently downbiased inbreeding levels.

Our results indicate that genetic rescue is likely rather short lived and that inbreeding in small natural populations is often more severe than detected from genetically verified pedigrees. To improve our understanding about the underlying mechanisms of genetic rescue, future studies should be directed at exploring the relationship between functional genes and inbreeding depression.