AGBT2023™
AGRICULTURE

March 27-29, 2023 • San Antonio, TX

Element Biosciences

Booth #302

Your Science Can't Wait

COMING SOON

Avidity Cloudbreak™

- Faster DNA sequencing run times
- Simplified workflow with linear library loading
- Industry leading quality
- Lowest run costs at $5/Gb

AGBT2023™ AGRICULTURE

March 27-29, 2023 • San Antonio, TX

We are honored to welcome you to the Advances in Genome Biology and Technology (AGBT) 2nd annual Ag meeting. You are among the world's leading genome researchers, data scientists, breeders, policy influencers, funders, and technology innovators from the global North and South embracing opportunities to redesign terrestrial and aquatic agriculture.

Over the next few days, we will focus on the integration of genomics and agriculture to address the escalating needs of a changing Earth.

AGBT Ag is the third global meeting organized by The Genome Partnership, a Saint Louis based non-profit that advances research, promotes education and expands commerce in genome science and technology. The Genome Partnership has been proudly organizing AGBT meetings since 1999.

Organizing Committee

We are incredibly grateful to Sarah Hearne (OC Chair) and the entire Organizing Committee for the countless hours they invested in making AGBT Ag23 a content-rich experience.

Over the next several days we will gather to collaborate, share the latest scientific discoveries, and discuss new technological breakthroughs thanks in large part to the work of our organizing committee.

Xiaofeng Cao
Principal Investigator-Institute of Genetics and Developmental Biology, Chinese Academy of Sciences (CAS)

Jack Dekkers
Distinguished professor, section leader of animal breeding and genetics, Iowa State University

Appolinaire Djikeng
Director, Centre for Tropical Livestock Genetics and Health, The University of Edinburgh

Sarah Hearne
Principal Scientist, International Maize and Wheat Improvement Center (CIMMYT)

Charlie Johnson
Executive Director Genomics & Bioinformatics, Texas A&M AgriLife

Renee Lafitte
Deputy Director for Crops R&D, Bill and Melinda Gates Foundation

Nathan Lakey
President and Chief Executive Officer, Orion Genomics

Daniela Lourenco
Associate Professor, University of Georgia

Susan McCouch
Professor of Plant Breeding and Genetics, Cornell University

Len Pennacchio
DOE Joint Genome Institute, Lawrence Berkeley National Laboratory

Alison Van Eenennaam
Animal Biotechnology and Genomics, Department of Animal Science, University of California, Davis

Bruce Walsh
Professor of Ecology and Evolutionary Biology, University of Arizona

Susan Wessler
Distinguished Professor of Genetics, University of California, Riverside

AGBT2023™
AGRICULTURE

We gratefully acknowledge the generous support provided by the following companies

GOLD

Element Biosciences

SILVER

CORTEVA™ agriscience

BRONZE

PacBio

PerkinElmer® For the Better

EXHIBITORS & CONTRIBUTING SPONSORS

AgriPlex GENOMICS

DAICEL arbor biosciences

BIO ECHO Life Sciences

Complete GENOMICS Powered By MGI

computomics® machine learning-based data analysis

METWARE BIO Innovative Metabolomics Insights For Better Health

mirxes

Oxford NANOPORE Technologies

NEOGEN®

PARAGON GENOMICS

seqWell

SINGULAR GENOMICS

sptlabtech

SYNBIO TECHNOLOGIES

TECAN.

ThermoFisher SCIENTIFIC

General Information

Registration / Hospitality Desk

Saturday, March 25th	2:00 PM – 7:00 PM
Sunday, March 26th	7:00 AM – 7:00 PM
Monday, March 27th	7:30 AM – 7:00 PM
Tuesday, March 28th	7:30 AM – 7:00 PM
Wednesday, March 29th	7:30 AM – 7:00 PM

Name Badges
Please wear your name badge & lanyard at all times. Security will refuse entry to anyone not associated with the AGBT meeting.

Transportation
AGBT complimentary shuttles are provided out of San Antonio International Airport (SAT).

Arrival Shuttles:
Sunday, March 25th: 1:00 p.m. - 8:00 p.m.
Departure Shuttles:
Thursday, March 30th: 4:00 a.m. - 2:00 p.m.

Social Events
Welcome Dinner – **Monday, March 27th 6:00 pm – 9:00 pm** on Aunt Mary's Lawn.

Click2Win Party – **Tuesday, March 28th beginning at 7:35 pm** in the Sponsor Promenade and Exhibit Hall. Grand Prize is complimentary registration to AGBT Ag 2024 and will be awarded on Thursday's Plenary Session. (This prize is non-transferable.)

Farewell Dinner Party – **Wednesday, March 29th 7:00 pm – 10:00 pm** at Henry's Hollow.

Conference Technology
Download our official AGBT conference mobile app for easy access to the agenda, abstracts, and more. Check your email for the link to download the app or visit our registration desk for assistance.

Start Tweeting!

@agbt #AGBTAg23 #AGBTAg

@agbt.org

@AGBT – Advances in Genome Biology and Technology

Visit our website: https://www.agbt.org/events/agricultural-meeting-overview/


WE PUSH PROGRESS FORWARD.
Our future-forward approach helps farmers – in Iowa and around the world - optimize their inputs, reduce their carbon footprint and preserve natural resources.
Enriching the lives of those who produce and those who consume for generations to come. Learn more at keepgrowing.corteva.com
KEEP GROWING.
CORTEVA™
agriscience
®, TM, SM Trademarks and service marks of DuPont, Dow AgroSciences or Pioneer, and their affiliated companies or their respective owners. ©2020

AGBT2023™
AGRICULTURE

Agenda

Saturday, March 25	
2:00 p.m. – 7:00 p.m.	**AGBT Meeting Registration**
Sunday, March 26	Transfers 1 p.m. – 8 p.m. San Antonio International Airport (SAT)
7:00 a.m. – 7:00 p.m.	**AGBT Meeting Registration** Hill Country Foyer
Monday, March 27	**Scientific Program**
7:30 a.m. – 7:00 p.m.	**Meeting Registration / Hospitality Desk** Hill Country Foyer
7:30 a.m. – 9:00 a.m.	**Breakfast** Courtyard Deck
9:00 a.m. – 9:10 a.m.	**Opening Remarks** Independence Ballroom 1-5
Plenary Session I:	**Global Opportunities and Challenges for Agriculture, I (**Sarah Hearne, Principal Scientist, International Maize and Wheat Improvement Center (CIMMYT), Session Chair)
9:10 a.m. – 9:40 a.m.	Lindiwe Majele Sibanda, Chair of the CGIAR System Board, Extraordinary Professor, University of Pretoria, South Africa, Council Chair, National University of Science and Technology, Zimbabwe “The imperative to transform agriculture in a climate crisis: a global challenge that must be met through partnerships”
9:40 a.m. – 10:10 a.m.	Mark Cooper, University of Queensland “Revisiting the breeder’s equation: Implications of our advancing understanding of trait genome-to-phenome relationships for predictive breeding”
10:10 a.m. – 10:30 a.m.	Rebecca Clarke (Abstract Selected) Agresearch “Investigation of genomic regions of interest in animal genomes using ONT adaptive sampling”
10:30 a.m. – 11:00 a.m.	**Coffee Break, Sponsors and Exhibits** Hill Country Ballroom & Sponsor Promenade

Plenary Session II:	**Global Opportunities and Challenges for Agriculture, II (**Nathan Lakey, Orion Genomics, Session Chair) Independence Ballroom 1-5
11:00 a.m. – 11:30 a.m.	Rob Martienssen, Cold Spring Harbor Laboratory “Germline RNA interference and the drive for maize domestication”
11:30 a.m. – 11:50 a.m.	Aayudh Das (Abstract Selected), Pennsylvania State University “The functional genomics of local adaptation in sorghum landraces”
11:50 a.m. – 12:10 p.m.	Nadia Kamal (Abstract Selected), Helmholtz Munich “A healthy crop like no other: the mosaic reference genome of hexaploid oat”
12:15 p.m. – 1:30 p.m.	**Lunch** Independence Lawn
Plenary Session III:	**Applying the Science, I** (Bruce Walsh, Professor of Ecology and Evolutionary Biology, University of Arizona, Session Chair) Independence Ballroom 1-5
1:30 p.m. – 2:00 p.m.	Alison Van Eenennaam, Cooperative Extension Specialist, Animal Genomics and Biotechnology, University of California, Davis “Outcomes from embryonic editing of livestock genomes”
2:00 p.m. – 2:30 p.m.	Leena Tripathi, International Institute of Tropical Agriculture (IITA), Kenya “The potential of genome editing for sustainable agriculture”
2:30 p.m. – 2:50 p.m.	Thomas Poorten (Abstract Selected), Pairwise “Elimination of pungency in allotetraploid Brassica juncea through gene editing of the multicopy myrosinase gene family”
2:50 p.m. – 3:20 p.m.	**Coffee Break, Sponsors and Exhibits** Hill Country Ballroom & Sponsor Promenade
3:25 p.m. – 3:45 p.m.	**Gold Sponsor Talk – Element Biosciences** Carlos Congrains “Genomics applied to species identification of fruit fly (Diptera: Tephritidae) pests” Independence Ballroom 1-5
3:45 p.m. – 4:00 p.m.	**Silver Sponsor Talk – Corteva** Gina Zastrow-Hayes & Heather Snowgren “Accelerating product development with genomics at Corteva Agriscience” Independence Ballroom 1-5

Plenary Session IV:	**Enabling Approaches and Informatics,** (Len Pennacchio, DOE Joint Genome Institute, Lawrence Berkeley National Laboratory, Session Chair) Independence Ballroom 1-5
4:00 p.m. – 4:30 p.m.	Robin Buell, University of Georgia "Use of complementary single cell omics to elucidate specialized metabolism in the medicinal plant, *Catharanthus roseus* (Madagscar periwinkle)"
4:30 p.m. – 5:00 p.m.	Farhad Hormozdiari, Google Health "Deep learning for genomics applications"
5:00 p.m. – 5:20 p.m.	John Lovell (Abstract Selected), HudsonAlpha Institute for Biotechnology "Integrating comparative and quantitative genetics to accelerate breeding efforts"
5:20 p.m. – 5:40 p.m.	AGBT's Next Gen Leadership Awards
6:00 p.m. – 9:00 p.m.	**Welcome Dinner** Aunt Mary's Lawn
Tuesday, March 28	**(Morning)**
6:30 a.m. – 7:30 a.m.	Rise and Shine with AGBT – Optional Activity Meet in Lobby
7:30 a.m. – 7:00 p.m.	**Hospitality/Information Desk** Hill Country Foyer
7:30 a.m. – 9:00 a.m.	**Breakfast** Courtyard Deck
Plenary Session V:	**Applying the Science, II (**Jack Dekkers, Distinguished Professor, section leader of animal breeding and genetics, Iowa State University, Session Chair**)** Independence Ballroom 1-5
9:00 a.m. – 9:30 a.m.	Anna Sonesson, Norwegian Food Research Institute (Nofima) "Genetics in aquaculture"
9:30 a.m. – 9:50 a.m.	Harry Lamb (Abstract Selected), Queensland Alliance for Agriculture and Food Innovation "On-farm genomics for rapid genotyping and pathogen monitoring in livestock"
9:50 a.m. – 10:10 a.m.	R. Kelly Dawe (Abstract Selected), University of Georgia "Synthetic maize centromeres transmit chromosomes across generations"

10:10 a.m. – 10:30 a.m.	Sergey Koren, (Abstract Selected) National Human Genome Research Institute (NHGRI) “Automated telomere-to-telomere crop genome assembly using only Oxford Nanopore sequencing”
10:30 a.m. – 11:00 a.m.	**Coffee Break, Sponsors and Exhibits** Hill Country Ballroom & Sponsor Promenade
Plenary Session VI:	**Informatics and Analytics Session VI:** (Daniella Lourenco, University of Georgia, Session Chair) Independence Ballroom 1-5
11:00 a.m. – 11:30 a.m.	Gregor Gorjanc, University of Edinburgh “Managing and improving populations in the era of mega-scale genomics”
11:30 a.m. – 12:00 p.m.	Gustavo de los Campos, Michigan State University “Sparse selection indices for high dimensional phenotypes, sparse genomic prediction, and both”
12:00 p.m. – 2:00 p.m.	**AGBT Lunch** Independence Lawn
12:45 p.m. – 2:30 p.m.	**Ag T2T Workshop Lunch** Sergey Koren, Wes Warren, Tim Smith, Ben Rosen, Zhigui Bao Independence Ballroom 1-5
Plenary Session VII:	**From FASTQ to Fork** (Alison Van Eenennaam, Cooperative Extension Specialist, Animal Genomics and Biotechnology, University of California, Davis, Session Chair) Independence Ballroom 1-5
3:00 p.m. – 3:20 p.m.	Hao Cheng (Abstract Selected), University of California, Davis “Homomorphic encryption to enable sharing of confidential data in agricultural genome to phenome”
3:20 p.m. – 3:32 p.m.	**Bronze I Talk – PacBio** Jake Brandenburg, PacBio “Ultra-High throughput multi-omics workflows for HiFi sequencing” Independence Ballroom 1-5
3:32 p.m. – 3:44 p.m.	**Bronze II Talk – PerkinElmer** Eva M. Farre, Ph.D., Associate Professor of Plant Biology, Michigan State University “Variation of diurnal transcriptomes across potato genotypes” Independence Ballroom 1-5
3:45 p.m. – 4:00 p.m.	**Poster Flash talks**

4:00 p.m. – 6:00 p.m.	**Poster Session & Exhibit Hall, Wine Reception** Hill Country Ballroom
6:00 p.m. – 7:30 p.m.	**Dinner** Independence Lawn
Beginning at 7:35 p.m.	**Click2Win Party** (Engage with sponsors and exhibitors to win great prizes, including free registration to next year's AGBT Ag)
Wednesday, March 29	**(Morning)**
6:30 a.m. – 7:30 a.m.	**Rise and Shine with AGBT – Optional Activity** Meet in Lobby
7:30 a.m. – 7:00 p.m.	**Hospitality/Information Desk** Hill Country Foyer
7:30 a.m. – 9:00 a.m.	**Breakfast** Courtyard Deck
Plenary Session VIII:	**Building From Basic Science,** (Appolinaire Djikeng, Director, Centre for Tropical Livestock Genetics and Health, The University of Edinburgh, Session Chair**)** Independence Ballroom 1-5
9:00 a.m. – 9:30 a.m.	Antonio Augusto Franco Garcia, University of Sao Paulo "Genetic mapping in autopolyploids"
9:30 a.m. – 10:00 a.m.	Jo Newton, Agriculture Victoria Research "Genomics: an era of innovation in dairy cattle breeding"
10:00 a.m. – 10:20 a.m.	Temitayo Olagunju, (Abstract Selected), University of Idaho "The first complete ungapped T2T Y-chromosome assemblies of cattle and sheep"
10:20 a.m. – 11:00 a.m.	**Coffee Break, Sponsors and Exhibits** Hill Country Ballroom & Sponsor Promenade
Plenary Session IX:	**Quantitative and Population Genetics: Tools and Applications,** (Charlie Johnson, Executive Director Genomics & Bioinformatics, Texas A&M AgriLife, Session Chair)
11:05 a.m. – 11:25 a.m.	Matias Bermann (Abstract Selected), University of Georgia "Single-step genome-wide association analysis with p-values for large genotyped populations"

11:25 a.m. – 11:45 a.m.	Klara Verbyla (Abstract Selected), Center for Aquaculture Technologies “Commercial implementation of genomic selection in Tasmanian Atlantic salmon: Scheme evolution, validation, and further developments”
11:45 a.m. – 12:05 p.m.	Keith Gardner (Abstract Selected), International Maize and Wheat Improvement Center (CIMMYT) “Estimating genetic gain in CGIAR breeding programs: challenges and opportunities”
12:10 p.m. – 1:30 p.m.	**Lunch** Independence Lawn
Plenary Session X:	
1:30 p.m. – 2:25 p.m.	**Harnessing Talent** (Daniela Lourenco, University of Georgia and Sarah Hearne Principal Scientist, International Maize and Wheat Improvement Center (CIMMYT), Session Chairs) Independence Ballroom 1-5
Panel Discussion **Interactive Session**	**Mentoring in Science** (Jo Newton, Agriculture Victoria Research, Facilitator) **Next Generation Scientists** (Daniela Lourenco, University of Georgia, Session Chair)
2:30 p.m. – 2:50 p.m.	Valentine Ntui (Abstract Selected), International Institute of Tropical Agriculture “Engineering banana for resistance to diseases through CRISPR activation”
2:50 p.m. – 3:20 p.m.	**Coffee Break** Windmill Plaza
Plenary Session X continued:	
3:20 p.m. – 3:40 p.m.	Chensong Chen (Abstract Selected), The University of Queensland “Machine learning for performance prediction in species with polyploid and hybrid genomes
3:40 p.m. – 4:00 p.m.	Katharine Jenike (Abstract Selected), Johns Hopkins University “Alignment-free and interactive pan-genome visualization reveals lineage specific sequence conservation in Solanum”
4:00 p.m. – 4:20 p.m.	Daniel Tolhurst (Abstract Selected), The Roslin Institute, University of Edinburgh “Genomic selection for broad and specific adaptation using regressions on known and latent environmental covariates”

Plenary Session XI:	**From the Cutting Edge-New Ideas and Opportunities,** (Renee Lafitte, Deputy Director for rops R&D, Bill and Melinda Gates Foundation, Session Chair) Independence Ballroom 1-5
4:25 p.m. – 4:45 p.m.	Rupashree Dass (Abstract Selected), Computomics “Machine learning-based phenotype prediction help in detecting significantly associated genetic markers”
4:45 p.m. – 5:05 p.m.	John Bergeron (Abstract Selected), University of Nevada “Where’s Waldo? Describing resilience linked land use behaviors of sheep via GPS collars”
5:05 p.m. – 5:15 p.m.	Closing Comments, and Click2Win Announcements, (Sarah Hearne, Principal Scientist, International Maize, and Wheat Improvement Center (CIMMYT) and AGBT Ag Conference Chair)
7:00 p.m. – 10:00 p.m.	**Farewell Dinner Party** Henry’s Hollow
Thursday, March 30	Transfers 4 a.m. – 2 p.m. to San Antonio International Airport (SAT)

PacBio

GENOMICS AT SCALE

BREEDING PROGRAMS

Capture variations to advance breeding

PEST + DISEASE RESEARCH

Protect crops and livestock from pests and disease

GENOME ENGINEERING

Improve reliability of gene editing workflows

FIND US AT AGBT AG

PACBIO WORKSHOP | BOOTH 309 | FREDERICKSBURG D

PACB.COM/AGBT-AG-23

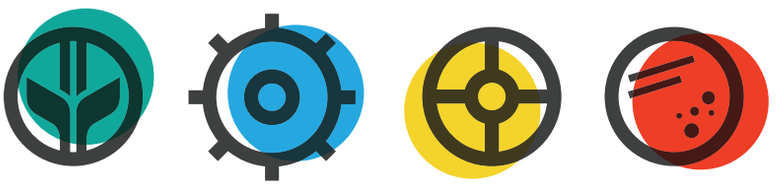
AGBT2024™
AGRICULTURE

APRIL 15-17, 2024
THE SHERATON AT WILD HORSE PASS
PHOENIX, AZ

LEARN MORE AGBT.ORG

Abstract Table of Contents

page



AGBT2023™

AGRICULTURE

OPTIMIZE YOUR NGS WORKFLOWS
AT AGBT AG 2023

Visit PerkinElmer in the Fredericksburg B Suite at AGBT Ag 2023 to Optimize Your NGS Workflows

Learn how PerkinElmer is eliminating the challenges associated with NGS by providing labs with complete, single-source solutions encompassing everything from sample to result. With application-based workflows, PerkinElmer increases your throughput while simplifying the generation of robust, reliable results.

These end-to-end solutions are designed to help you escape from your problems – from improving nucleic acid yields, throughput and reliability to reducing costs and bias, ***PerkinElmer has the answer you need.***

JOIN OUR SPONSOR WORKSHOP:

Variation of Diurnal Transcriptomes across Potato Genotypes
Eva Farre, Ph.D. Michigan State University

Tuesday March 28, 2023 | 3:32 pm – 3:44 pm
Independence Ballroom 1-5

For research use only. Not for use in diagnostic procedures.

PerkinElmer®

Basic Science: Genomes to a Reference

Haplotype resolved genome assembly of potato dihaploids capturing North American allelic diversity of tetraploid parents

Brieanne Vaillancourt[1], Kathrine Mailloux[1], John P. Hamilton[1,2], Xiaoxi Meng[3], Paul C. Bethke[4], David Douches[5], Shelley Jansky[4], Ek Han Tan[6], Laura Shannon[3], Jeffrey B. Endelman[7], and C. Robin Buell[1,2,8]

1 Center for Applied Genetic Technologies, University of Georgia, Athens, GA 30602, USA
2 Department of Crop & Soil Sciences, University of Georgia, Athens, GA 30602, USA
3 Department of Horticultural Science, University of Minnesota, St. Paul, MN 55108, USA
4 USDA Vegetable Crops Research Unit and University of Wisconsin-Madison, Madison 53706, WI, USA
5 Department of Plant, Soil and, Microbial Sciences, Michigan State University, East Lansing, MI 48824, USA
6 School of Biology and Ecology, University of Maine, Orono, ME 04469, USA
7 Department of Horticulture, University of Wisconsin-Madison, Madison, WI 53706, USA
8 Institute of Plant Breeding, Genetics & Genomics, University of Georgia, Athens, GA 30602, USA

Potato (Solanum tuberosum) is the third most important food crop in the world with a value of over $3.9 billion in the United States. Cultivated potato, Solanum tuberosum, is a highly heterozygous tetraploid (2n = 4x = 48) and as a consequence, breeding is primarily phenotypic-based with the breeder selecting an F1 from a cross of two tetraploid parents. Diploid (2n = 2x = 24) breeding of inbred/F1 hybrids has the potential to revolutionize potato breeding due to the simpler genetics and the ability to use true seed instead of seed tubers. Primary dihaploids for this project were created by crossing a tetraploid female parent with pollen from haploid inducer male parent IVP101 or IVP48 (Solanum tuberosum group phureja). A set of 91 female fertile dihaploids were checked for aneuploidy using Illumina gDNA sequencing data. Based on aneuploidy results, vigor, and diversity, a set of 20 female-fertile dihaploids were selected for chromosome-scale long read assembly using PacBio HiFi reads. Hifiasm was used to assemble the version 1 phased genome assemblies and Ragtag with the DMv6.1 reference genome assembly was used to scaffold the assemblies into 24 chromosomes. Version 2 assemblies will be refined using Hi-C contacts with Juicer and 3D-DNA. Nanopore full-length cDNA sequencing of young leaf and tuber libraries were used to annotate the assemblies. This project will provide high quality contiguous assemblies for 20 dihaploids facilitating the understanding of the allelic diversity of fertile diploid North American germplasm, a key component of converting US potato breeding to a diploid/F1 hybrid approach.

Deciphering the mechanisms of duckweed sexual reproduction

Cristian Mateo-Elizalde[1], Evan Ernst[1] and Rob Martienssen[1]

1 Howard Hughes Medical Institute, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, USA

Duckweeds are known as the smallest flowering plants, although many clones barely produce flowers either in nature or in laboratory conditions. This was one of the reasons that left Lemna behind with the advent of the molecular genetic era during the 1990's. Nonetheless, duckweeds have attracted the attention of investigators in applied sectors, some of them related to seek for an alternative source of food. Duckweeds, as aquatic plants, do not compete for agricultural land and their protein content in some species could raise until 40% of dry weight, becoming a good alternative for protein source (Rusoff et al., 1980). However, there are still many physiological and molecular aspects not well explored in Lemnnaceae.

Recent advances in high-throughput genomic technologies have opened up an interesting opportunity to unveil duckweed molecular biology raising the number of duckweed publications exponentially over the last 20 years, and we are now in what many authors has defined as "The return of the Lemnaceae" (Acosta et al., 2021).

Because of their predominant clonal growth habit, sexual reproduction is still one of the least studied aspects in this plant family. The plant hormone, salicylic acid, a floral inducer in duckweed, has been used for transcriptomic and physiological studies (Fouronjian et al., 2021), along with the identification of key flowering regulatory genes such as LgFT1 and LaFTL1 (Fu et al., 2020; Yoshida et al., 2021). However, the mechanism of SA induced flowering, and how is it connected to flowering in nature, is still unknown.

Here, we present our transcriptomic analysis conducted on duckweed pistils, anthers, seeds and cotyledons of the short-day plant L. aequinoctialis with the aim of providing the first specific and global characterization of the duckweed transcriptome related to flowering. Our study contributes to the further characterization flowering development, and reveals evidence of alternative strategies in gene expression control between sexual and clonal propagation, shedding light on this unexplored area of plant physiology and development.

(continue to page 22)

(continued from page 21)

Acosta K, Appenroth KJ, Borisjuk L, et al. Return of the Lemnaceae: duckweed as a model plant system in the genomics and postgenomics era. Plant Cell. 2021;33(10):3207-3234. doi:10.1093/plcell/koab189

Fourounjian P, Slovin J, Messing J. Flowering and Seed Production across the Lemnaceae. Int J Mol Sci. 2021;22(5):2733. Published 2021 Mar 8. doi:10.3390/ijms22052733

Fu L, Tan D, Sun X, Ding Z, Zhang J. Transcriptional analysis reveals potential genes and regulatory networks involved in salicylic acid-induced flowering in duckweed (Lemna gibba). Plant Physiol Biochem. 2020;155:512-522. doi:10.1016/j.plaphy.2020.08.001

Rusoff L L, Blakeney E W, Culley D D. Duckweeds (Lemnaceae family): a potential source of protein and amino acids. J Agric Food Chem. 1980;28;848–850. doi: 10.1021/jf60230a040

The first complete ungapped telomere-to-telomere (T2T) Y-chromosome assemblies of cattle and sheep

Temitayo Olagunju[1], Benjamin Rosen[2], Timothy Smith[3], Tracy Hadfield[4], Sergey Koren[5], Holly Neibergs[6], Noelle Cockett[4], Brenda Murdoch[1]

1 University of Idaho, Animal, Veterinary and Food Sciences (AVFS), Moscow, Idaho, U.S.A.
2 USDA, ARS, Animal Genomics and Improvement Laboratory, Beltsville, Maryland, U.S.A.
3 USDA, ARS, U.S. Meat Animal Research Center, Clay Center, Nebraska, U.S.A.
4 Utah State University, Animal, Dairy, & Veterinary Sciences (ADVS), Logan, Utah, U.S.A.
5 National Human Genome Research Institute, NIH, Genome Informatics Section, Bethesda, Maryland, U.S.A.
6 Washington State University, Animal Sciences, Pullman, Washington, U.S.A.

The Y-chromosome of mammals contains the majority of genes that are critical for sex determination and play vital roles in spermatogenesis and male fertility. It has been difficult to assemble due to the high number of repeats and palindromic sequences spanning more than half of its length. Even the human Y-chromosome was less than half complete until recently. In the Bovidae family, there is no species yet with a complete Y-chromosome; the most complete assemblies are found in cattle and sheep, spanning 16Mb and 10.6Mb, respectively. This information gap results in a poor understanding of the structure and variation of this vital chromosome.

We present the first complete ungapped Y-chromosomes of cattle and sheep from draft telomere-to-telomere (T2T) assemblies. The chromosome-level assemblies are 59Mb and 25Mb long for cattle and sheep, respectively. With de-novo and reference-guided annotations, we identified and characterized the pseudo-autosomal region (PAR), male-specific Y (MSY), X-degenerate, and the ampliconic regions. Our T2T assemblies have captured the genomic structures contained in the most up-to-date assemblies of both the cattle and sheep Y-chromosomes. A cross-comparison of the two species revealed the structural differences and similarities between these two members of the Bovidae family.

This research provides the most complete information on the structure and organization of the previously unknown highly repetitive heterochromatic and ampliconic regions of these two members of the Bovidae family. The findings in this work open new opportunities for further studies to explore the genomic architecture and variation of cattle and sheep Y-chromosomes.

Assemblage and annotation of microbial MAGs from secondary forest land in Sarawak, Malaysia: To plant or not to plant

Zahidah Ayob[1], Nor Azizah Kusai[1], Nur Maisarah Jantan[1], Vijaya Subramaniam[1], Ahmad Afandi Murdi[2], Hasimah Mos[2], Meilina Ong-Abdullah[2], Idris Abu Seman[3]

1 Malaysian Palm Oil Board, Peat Ecosystem and Biodiversity Unit, Kajang, Selangor, Malaysia
2 Malaysian Palm Oil Board, Agronomy and Geospatial Technology Unit, Kajang, Selangor, Malaysia
3 Malaysian Palm Oil Board, Plant Pathology and Biosecurity Unit, Kajang, Selangor, Malaysia

The decision to convert secondary forests to oil palm plantations benefits from having the knowledge of its soil microbial composition and genomic potential. To date, comprehensive metagenomic data generated from forest land in Borneo Malaysia remains scarce. In this study, we generated a de novo metagenome assembly on soil collected from a secondary forest in Sarawak, Malaysia. Metagenomic binning approach was applied to the dataset, leading to the recovery of at least 19 metagenome-assembled-genomes (MAGs) with more than 50% completeness and less than 5% contamination. Only one out of the 19 MAGs was of archaeal-origin. Several carbohydrates, nitrogen and methane metabolism genes were identified from the MAG annotation highlighting the intrinsic genomic potential of the microbes in biochemical cycling. This study provides a preliminary framework and serves as a new metrics to facilitate the selection of secondary forest land that is optimal for oil palm plantation while being mindful of preserving and conserving the microbial diversity.

Bioinformatics & Data Science

BEAKER (Biological Embeddings Acquired from KmERs), a novel model for species identification

Rachael Ashby[1,2], Joseph Guhlin[1,3], Marty Faville[4], Peter Dearden[1,3], Shannon Clarke[1,2]

1 Genomics Aotearoa, www.genomics-aotearoa.org.nz, New Zealand.
2 AgResearch, Invermay, Mosgiel, New Zealand.
3 University of Otago, Department of Biochemistry, Dunedin, New Zealand.
4 AgResearch, Grasslands, Palmerston North, New Zealand.

The boom in sequencing technologies has made access to genomic sequencing readily available with multiple applications of sequencing now commonplace across biology. This has increased our knowledge of biology and biological systems, as well as enabled the study of many species. However, DNA extractions are not species specific and often other species will be sequenced along with the species of interest. This has led to the need for new methods to identify these 'Hitchhiker' species to allow us to move from single species analyses to understanding the ecosystem of a species. Advances in Natural Language Processing (NLP) methods have led to these being implemented in genomics. Originally used to process language, we have applied these methods to DNA kmers to implement new models for separating sequences. Here we propose a novel method for embedding kmers that differs from other implementations by first training the embeddings to calculate the pairwise alignment distance between any two kmers. This allows a fixed input size of embeddings regardless of the length of kmers. Further, we utilise a novel pre-training approach by working to identify sequences from the same organism, recognize reverse complement sequences, and to identify erroneous kmers. Using a two-tiered approach to creating kmer embeddings allows a new model that anyone can utilise and adapt to their requirements as well as enabling our model to be trained on very large databases such as NCBI's nt with our tripleloss formula. These embeddings can then be used in downstream machine learning tasks. The applications of this model are limitless, with development utilising datasets including whole genome sequencing, genotyping-by-sequencing, transcriptomic, and eDNA to name a few. Furthermore, the methodology is not specific to any sequencing technology, Examples including genotyping-by-sequencing of perennial ryegrass (Lolium perenne), which frequently includes sequence from a symbiotic endophytic fungus (Epichloë festucae var. lolii), will be presented to demonstrate the model.

The Practical Haplotype Graph: A graph-based approach for imputing genotypes

Ana Berthel[1], Peter Bradbury[1,2], Terry Casstevens[1], Bethany F. Econopouly[1], Evan Long[3], Lynn C. Johnson[1], Zachary R. Miller[1], Brandon Monier[1], Mª Cinta Romay[1], Baoxing Song[1], Matthew Wiese[1], Edward S. Buckler[1 2 3]

1 Cornell University, Institute for Genomic Diversity, Ithaca, NY, USA
2 USDA-ARS, Ithaca, NY, USA
3 Cornell University, Plant Breeding and Genetics Section, Ithaca, NY, USA

The Practical Haplotype Graph (PHG) is a computational framework that uses a pangenome approach to impute genotypes from low-coverage sequence data. The PHG represents the pangenome as a trellis graph composed of genic and intergenic regions to capture diversity across germplasm and efficiently store genomic information. Genome assemblies and/or moderate to high coverage whole genome sequencing data are used to populate the graph with haplotypes. These haplotypes can then be clustered into a set of consensus haplotypes for each region.

Once a graph has been built, the PHG can be used to infer genotypes from low-coverage data. Sequences for a given sample are aligned to the pangenome to identify haplotypes matching each entry, and a Hidden Markov Model (HMM) finds the most likely path through the graph. Information from the path can be used to impute variants, call rare alleles, or construct custom genomes.

With this framework, only representative founder genotypes in a breeding program must be sequenced at a high coverage for genome-wide prediction. Progeny can be sequenced using low-cost, low-coverage technologies, and the PHG accurately infers their genotypes. This allows a larger number of samples to be processed, so the population sizes in a breeding program can be increased. The PHG is compatible with data from multiple technologies including target genotyping, reduced genome representation, and skim sequencing, and is generalizable to different species of interest. We are currently deploying the PHG in three species—maize, cassava, and sorghum—and will present progress on these PHGs here.

The PHG is stored as either a Postgres or SQLite database, and users can interact with it through a variety of interfaces. In addition to the PHG's native API, we have developed an R package called rPHG that uses common Bioconductor classes, data structures, and methods, and facilitates downstream analysis and integration with other packages. We have also built a web service that implements Breeding API (BrAPI) genotyping calls for easy retrieval of imputed haplotype information. Using the PHG and its interfaces, breeders have access to valuable genomic information without the need for large-scale, high-coverage sequencing.

AgriSeqRI 1.0: reporting utility for the Canine Traits and Disorders Panel

Srinivas Chadaram[1], Angela Burrell[1], Krishna Reddy Gujjula[1], Nikhil Anjuri[2], Anuradha Udumudi[2], Srinivas Udumudi[2] and Haktan Suren[1]

1 Thermo Fisher Scientific, Austin, TX,
2 ATS GeneTech Pvt Ltd, India

The AgriSeqRI is a bulk report generation plugin primarily developed for the AgriSeq™ Canine Traits and Disorders panel run using the Torrent Suite™ Software (TSS). End users can generate individual reports for each sample on a sequencing run using AgriSeqRI.

AgriSeqRI can be run manually or automatically (via plan run) after a sequencing run is completed on Ion Torrent. The plugin retrieves the genotype data from Torrent Variant Caller (TVC) and AgriSum Toolkit and outputs a comprehensive report for each sample for all 154 markers (133 disorders, 21 traits associated). The report includes sample, reference, panel (test) details, and a summary of the results. In addition, information including the condition, genotype (reference and variant allele), inheritance type, and OMIA reference, is provided for each marker. A brief explanation of the genetic trait, breed specificity, genetics and inheritance gathered from several publicly available publications are also reported.

AgriSeqRI allows users to export the reports in bulk mode per customer for ease of processing. Supported output formats include JSON and PDF. The report can be fully customized allowing users to add or /remove the report section and even facilitating custom logos on the final report.

AgriSeqRI enables veterinarians, and veterinary service labs to enhance their product offerings related to the AgriSeq™ Canine Traits and Disorders panel by automating the report creation process. AgriSeqRI is publicly available and can be downloaded via Thermo Fisher Cloud and installed on TSS at no additional cost.

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

Enrichment of the groundwater chemical composition data for convergence informatics

Alexey V. Gulyuk[1], Akhlak Mahmood[1], Paul Westerhoff[2], and Yaroslava Yingling[1]

1 North Carolina State University, Materials Science and Engineering Department, Raleigh, NC, USA.
2 Arizona State University, Civil and Environmental Engineering Department, Tempe, AZ, USA.

The problem of reusing available natural resources and, particularly, removing various chemical pollutants from water sources is a topic that currently receives a lot of attention. Utilization of various agents (nanoparticles, hydrogels, or solvents) opens new paths to recovery the dangerous wastewater pollutants like hydrogen, phosphorus, or heavy metals.

The very first step for developing water cleaning strategies requires assessing full chemical composition of a target water source. In practice, extensive water quality analysis involves analyzing enormous amounts of data and requires rigorous data preparation and several preprocessing steps.

Here we want to present our vision of how the combination of experimental and ML-derived data can help to facilitate the increased accuracy of data analysis for further advances in clean water sustainability. Particularly, we focus on assembly, analysis, and completion of the water chemical composition dataset, which became one of cornerstones of the project. Data enrichment tools utilized in this work rely on Machine Learning algorithms and enable elements of Convergence Informatics, thus advancing data-driven research with the end goal of finding the most effective water treatment compounds and agents.

Breeders' genomic environment: a hub full of tools for analyzing genomic data

Lynn Johnson[1], Francisco J. Agosto-Perez[2], Peter Bradbury[1,3], Terry Casstevens[1], Bethany Econopouly[1], Zack Miller[1], Brandon Monier[1], Jeffry Morse[1], Matthew Wiese[1], Edward S Buckler[1,2,3]

1 Cornell University, Institute Biotechnology and Life Science, Ithaca, New York, USA
2 Cornell University, Section of Plant Breeding and Genetics, Ithaca, New York, USA
3 USDA-ARS, Ithaca, NY, USA 14853

Fast, reproducible, and robust bioinformatic pipelines are fundamental to the analysis of increasing amounts of genomic data. The Breeders' Genomics Hub is a JupyterHub instance setup with a collection of open-source tools used to collect, analyze and predict biological patterns based on phenotypic and genotypic data. It provides a single platform for data analysis and pipeline documentation. Utilization of Jupyter Notebooks within the hub promotes sharing and reproducibility of results.

One of the strengths of the Breeders' Genomics Hub is it provides researchers with access to powerful, yet user-friendly, data sources and software tools. It leverages the Breeding Application Programming Interface (BrAPI) web services to retrieve data from common breeding databases using BreeDBase, BMS and/or Gigwa. For example, rTASSEL and rPHG provide users with access to TASSEL and the Practical Haplotype Graph (PHG), a graph-based software tool for calling haplotypes from high-throughput sequencing data via R-front ends installed on the Hub. BioKotlin, a bioinformatics library offering the performance of compiled languages (>50-fold performance) with ease of Python-like scripting, is also available on the Hub. Using these tools, users may retrieve data from the PHG via rPHG, integrate the data into a single data structure with rTASSEL and BioKotlin, pre-process and visualize the data with common libraries, and apply machine learning or standard R models to make genomic predictions.

By using the Breeders' Genomics Hub to link data from multiple sources and perform advanced analysis plant breeders can gain valuable insights into the relationships between genotype and phenotype and make informed breeding decisions. The Breeders' Genomic Hub provides a user-friendly means of collaboration between scientists, easy integration with genomic analysis tools, and a convenient place to store data pipelines and results.

A healthy crop like no other: the mosaic reference genome of hexaploid oat

1 Helmholtz Zentrum München, German Research Center for Environmental Health, Plant Genome and Systems Biology, Ingolstädter Landstr. 1, 85764 Neuherberg, Germany, helmholtz-munich.de, nadia.kamal@helmholtz-muenchen.de
2 Lund University, ScanOats, Department of Chemistry, Pure and Applied Biochemistry, Box 124, 221 00, Lund, Sweden, nikos.tsardakas_renhuldt@tbiokem.lth.se
3 Lund University, ScanOats, Department of Chemistry, Pure and Applied Biochemistry, Box 124, 221 00, Lund, Sweden, johan.bentzer@tbiokem.lth.se
4 Helmholtz Zentrum München, German Research Center for Environmental Health, Plant Genome and Systems Biology, Ingolstädter Landstr. 1, 85764 Neuherberg, Germany, helmholtz-munich.de, h.gundlach@helmholtz-muenchen.de
5 Helmholtz Zentrum München, German Research Center for Environmental Health, Plant Genome and Systems Biology, Ingolstädter Landstr. 1, 85764 Neuherberg, Germany, helmholtz-munich.de, g.haberer@helmholtz-muenchen.de
6 Australian Research Council Centre of Excellence for Innovations in Peptide and Protein Science, Edith Cowan University, School of Science, 270 Joondalup Dr, Joondalup WA 6027, Australia, a.juhasz@ecu.edu.au
7 Helmholtz Zentrum München, German Research Center for Environmental Health, Plant Genome and Systems Biology, Ingolstädter Landstr. 1, 85764 Neuherberg, Germany, helmholtz-munich.de, thomas.lux@helmholtz-muenchen.de
8 Australian Research Council Centre of Excellence for Innovations in Peptide and Protein Science, Edith Cowan University, School of Science, 270 Joondalup Dr, Joondalup WA 6027, Australia, Commonwealth Scientific and Industrial Research Organisation, Agriculture and Food, 306 Carmody Road, St Lucia, QLD 4067, Australia, u.bose@ecu.edu.au
9 Walter and Eliza Hall Institute of Medical Research, Immunology Division, 1G Royal Parade, Parkville, VIC 3052, Australia, The Royal Melbourne Hospital, Gastroenterology, Grattan Street, Parkville, VIC 3052, Australia, tyedin@wehi.edu.au
10 Agriculture and Agri-Food Canada, Ottawa Research and Development Centre, 960 Carling Ave., C.E. Farm, Ottawa, ON K1A 0C6, Canada, Wubishet.Bekele@AGR.GC.CA
11 Commonwealth Scientific and Industrial Research Organisation, Agriculture and Food, 306 Carmody Road, St Lucia, QLD 4067, Australia, m.colgrave@ecu.edu.au
12 Brigham Young University, Plant & Wildlife Sciences Department, 4105 LSB, Provo, Utah, 84602, USA, rick_jellen@byu.edu
13 Brigham Young University, Plant & Wildlife Sciences Department, 4105 LSB, Provo, Utah, 84602, USA, jeff_maughan@byu.edu

(continue to page 31)

(continued from page 30)

14 Agriculture and Agri-Food Canada, Ottawa Research and Development Centre, 960 Carling Ave., C.E. Farm, Ottawa, ON K1A 0C6, Canada, Nick.Tinker@agr.gc.ca
15 Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Domestication Genomics, Corrensstrasse 3, 06466 Seeland, Germany, mascher@ipk-gatersleben.de
16 CropTailor AB, c/o Lund University, Department of Chemistry, Division of Pure and Applied Biochemistry, Box 124, 221 00, Lund, Sweden, olof.olsson@tbiokem.lth.se
17 Helmholtz Zentrum München, German Research Center for Environmental Health, Plant Genome and Systems Biology, Ingolstädter Landstr. 1, 85764 Neuherberg, Germany, helmholtz-munich.de, manuel.spannagl@helmholtz-muenchen.de
18 CropTailor AB, c/o Lund University, Department of Chemistry, Division of Pure and Applied Biochemistry, Box 124, 221 00, Lund, Sweden, nick.sirijovski@tbiokem.lth.se

Oat (Avena sativa L.) is an emerging crop with substantial benefits for human health. Oats are highly nutritious and have the potential to provide a plant-based dairy-free protein source. Whole grain oats are high in protein, polyunsaturated fatty acids, antioxidants, polyphenolic compounds, vitamins, minerals, and soluble fibres such as ß-glucans. ß-glucans are known for their beneficial health effects on the cardiovascular system and digestion, and for lowering blood cholesterol and the post-meal glycaemic index. Oats are generally considered a safe food source for patients with celiac disease. Moreover, oat has a low carbon footprint and cultivation requires fewer treatments with insecticides, fungicides or fertilizers.

However, due to the very large genome and its complexity (hexaploid, high repeat content, highly similar subgenomes), a well-characterised reference genome sequence for oat was not available for a long time. Here we present a fully annotated high-quality reference genome assembly of common oat from the Swedish Spring oat cultivar “Sang” and provide an in-depth investigation of the genomic factors underlying the unparalleled qualities of oats in human health and nutrition. Among others, we decoded the large gene family of cellulose synthase genes and provide a genetic foundation underlying the biosynthesis of oat ß-glucans and the health claims associated with blood cholesterol and post-meal glycaemic responses. We identified large-scale chromosome rearrangements in the oat genome by comparative structural analysis of different Avena species and traced them back to the polyploidization history of oat. This provides evidence for breeding barriers that were hidden in oat so far and caused difficulties in introgression breeding in oat. We analysed gene and repeat content as well as expression abundance across the three subgenomes of hexaploid oat and tested observed differences in relation to the structural rearrangements.

Exploring the role of cattle immunoglobulin loci in vaccine efficacy: insights from the rapid evolution of adaptive immunity in cows

Yana Safonova

Johns Hopkins University

An important challenge in vaccine development is to figure out why a vaccine succeeds in some individuals and fails in others. Antibody repertoires and immunoglobulin loci encoding them hold the key to answering this question. While previous personalized immunosequencing studies have successfully linked variations in human immunoglobulin loci to disease and vaccine efficacy, the use of computational tools for studying non-human species has been limited due to the high diversity and extreme repetitiveness of immunoglobulin loci, which makes them difficult to assemble and genotype. However, recent advances in full-length genome assembly have made it possible to analyze nearly complete sequences of immunoglobulin loci in vertebrate species. The analysis of the genomes of various cattle breeds has uncovered a unique structure of cattle immunoglobulin loci, which consist of three long tandem repeats. This structure allows for rapid evolution and results in a vast diversity of immunoglobulin genes across different cattle breeds. By performing an extended immunogenomics analysis of cattle immunoglobulin loci and combining this data with repertoire sequencing data, we were able to shed light on the variability of vaccine-induced antibody responses to bovine respiratory disease. Results of this analysis may aid in the development of more effective vaccines and improvement of existing breeding programs.

Breeding API (BrAPI): a standard API specification for plant breeding data

Peter Selby[1]

1 Cornell University, College of Agriculture and Life Sciences, School of Integrative Plant Science, Ithaca, New York, USA

Modern plant breeding research requires a large amount of data to function effectively. Data repositories are improving in their ability to store this data, but there is a growing need for interoperability between disparate data sources and applications. The Breeding Application Programming Interface (BrAPI) project offers a solution to this problem with a standardized RESTful web service API specification. This specification provides a standard data model for the plant breeding domain, plus a well-defined set of methods for interacting with the data. The goal of the project is to promote interoperability, data sharing, and open source code sharing across organizations who produce and consume data in this domain. The BrAPI project is a community built project and that community is well established and continuously growing. The standard is built based on concrete use cases to solve real interoperability challenges faced by the community. Beyond the core standard, the community has built a variety of open source tools and resources to help build and test implementations of the specification. The community is also constantly producing new BrAPI compliant applications, analysis tools, and visualizations that will work with any BrAPI compliant data source.

This presentation is intended for people in the overlap between plant scientists and research data management. It may be interesting to anyone interested in APIs, data sharing strategies, and collaborative software development.

Epigenetics & Epigenomics

Genomewide analysis of translational regulation across bovine tissues

Anna K. Goldkamp[1], Darren E. Hagen[1]

1. Dept of Animal and Food Sciences, Oklahoma State University, Stillwater, OK

Gene transcription and protein translation are core components in the process of gene expression. Previous research on the regulation of gene expression largely focuses on events prior to translation, including epigenetic regulation, transcription, and RNA processing. However, translation acts as an additional layer of regulation that plays an important role in gene expression and function. Highly expressed genes are thought to be codon-biased to support efficient translation, in which the encoded codons correspond to highly abundant tRNAs. Further, synonymous SNPs were once considered to be silent due to the degeneracy of the genetic code. However, synonymous SNPs may disturb protein abundance and function through alterations in translational efficiency and suboptimal pairing to lowly abundant tRNAs. Our previous study identified variation in tRNA expression within bovine liver and muscle tissue, suggesting tissue-specific modulation of translation. In addition, advances in sequencing technologies have recently permitted the study of the translatome, which refers to the entire population of mRNA associated with ribosomes for protein synthesis and can be investigated through ribosome profiling.

Here, we applied Quantitative Mature tRNA sequencing (QuantM-tRNA-seq) and ribosome profiling in order to investigate the relationship between tRNA expression and translational stalling events. Moreover, we have identified translationally regulated genes underlying tissue-specific biological processes and found that many upregulated and downregulated genes coincided with high and low translational efficiency respectively. We have also successfully defined stalling sites that depict the regulatory information encoded within the coding sequence of transcripts, which could control translation rate and facilitate proper protein folding. This work offers an atlas of distinctive stalling sites across bovine tissues, which provides an opportunity to predict codon optimality and understand tissue-specific mechanisms of regulating protein synthesis.

The dynamic epigenomic landscape of synthetic wheat

Manuel Spannagl, Maxim Messerer, Yuxuan Lan, Nadia Kamal, Klaus FX Mayer, Anthony Hall, Michael Bevan

Helmholtz Center Munich

The fusion of independent genomes to form hybrids and polyploids is common in plant evolution and breeding, and often leads to new gene expression patterns and traits. Epigenetic and chromatin-level changes can strongly influence gene expression levels and may contribute to new gene expression patterns, but this is currently poorly understood. Wheat is an ideal organism to study such changes, as the three component genomes in hexaploid wheat are highly similar and do not recombine, providing a stable platform to study homoeologs in different ploidy states. In a set of hexaploid wheat (AABBDD) synthetic lines derived from a common elite Triticum durum (AABB) and two genetically divergent Aegilops tauschii (DD) lines, we generated a comprehensive set of gene expression and epigenomic data. Different patterns of gene expression changes were seen in independent synthetic lines. Methylation patterns were highly conserved, with only ca 100 different genes showing altered methylation states that influence gene expression. This may represent a background state of methylation change. In contrast, relatively frequent changes in histone H3K27Me3 and H3K4Me3 marks were observed, with negative correlations between H3K27Me3, gene expression and accessible chromatin, and positive correlations between H3K4Me3 marks and accessible chromatin. Changes in the DD epigenome were more frequent than those in the AABB genome in hexaploid formation, with major changes in H3K27Me3 and H3K4me3 marks. Interestingly bivalent peaks containing both H3K4Me3 and H3K27Me3 were the least changeable, and were most likely to be resolved into loci without H3K27Me3 and those with only H3K4Me3 peaks. The consequences of these changes on the expression of conserved triad genes, unique genes and developmentally regulated genes are being explored.

DNA methylation identification and comparison from Oxford Nanopore Technologies (ONT), Pacific Biosciences (PacBio), and whole-genome bisulfate sequencing (WGBS) sequencing data

Brenda Murdoch[1], Shangqian Xie[1], Suraj Bhattarai[2], Stephanie McKay[2], Holly Neibergs[3], Tracy Hadfield[4], Noelle Cockett[4], Benjamin Rosen[5], Timothy Smith[6]

1 University of Idaho, Animal, Veterinary and Food Sciences, Moscow, Idaho, U.S.A.
2 University of Vermont, Animal and Veterinary Sciences, Burlington, Vermont, U.S.A
3 Washington State University, Animal Sciences, Pullman, Washington, U.S.A.
4 Utah State University, Animal, Dairy and Veterinary Sciences, Logan, Utah, U.S.A.
5 USDA, ARS, U.S. Meat Animal Research Center, Clay Center, Nebraska, U.S.A.
6 USDA, ARS, Animal Genomics and Improvement Laboratory, Beltsville, Maryland, U.S.A.

Cytosine methylation of DNA is a well-known epigenetic modification and has been implicated in various regulatory mechanisms in vertebrates. Whole genome bisulfite sequencing (WGBS) has been the most widely used technique for single base resolution of DNA methylation. Recent advances in long read sequencing technologies from Oxford Nanopore Technologies (ONT) and Pacific BioSciences (PacBio) have supported identification of modified bases in DNA, including the 5-methyl cytosine modification of CpG dinucleotide motifs (5mC) in vertebrates. Although the algorithms were developed to recognize 5mC in DNA sequence using human or model datasets, it is not clear how they perform with other species like livestock genome datasets. In this study, we identified and compared the DNA methylation from three platforms in sheep (ONT: 100X, PacBio: 70X, WGBS: 80X) and cattle (ONT: 125X, PacBio: 100X, WGBS: 80X). The results revealed that (1) A total of 43,865,765, 43,761,343, and 38,389,181 methylation sites were identified in sheep, and 48,445,027, 48,778,225 and 44,023,408 were identified in cattle, from ONT, PacBio and WGBS respectively. Long reads sequencing data from ONT and PacBio both identified more methylation sites than that from short reads sequencing data (WGBS), but WGBS identified higher consistent methylation sites (99.74%) than ONT (98.98%) and PacBio (98.99%). (2) The whole-genome density of methylated C sites (5mC/Total C) of ONT (0.0683 and 0.0709) and PacBio (0.0681 and 0.0713) were higher than that from WGBS (0.0598 and 0.0644) in both sheep and cattle, respectively. (3) The enrichment pattern of methylated CpG regions were broadly similar for all three platforms, and several enriched methylated sites were strongly associated with tandem repeats. However, some differences were observed in the WGBS, possibly due to read mapping limitations of short read data. These comparative results suggest that the human-trained algorithms are generally sufficient for 5mC identification in livestock species and illustrate that although global methylation patterns are similar among ONT, PacBio and WGBS, there are region-specific differences, especially for the repetitive regions. Mention of trade, product or corporation names in this abstract is for information only and does not imply endorsement or approval of the United States Department of Agriculture.

From FASTQ to Fork: Genomics - Assisted Breeding

Comparative transcriptome analysis reveals cell-wall–modifying genes associated with firmness in boiled storage roots of sweet potato (Ipomoea batatas)

Linly Banda[1,8]*, Bernice Waweru[2], Jean-Baka Domelevo Entfellner[2], Mukani Moyo[3], Arnold Onyango[4], Esther Magiri[5], Dorcus C. Gemenet[6], and Tawanda Muzhingi[7]

1 National University of Science and Technology, Department of Applied Biology and Biochemistry, P.O. Box AC 939, Ascot, Bulawayo, Zimbabwe
2 Biosciences Eastern and Central Africa-International Livestock Research Institute (BecA-ILRI) Hub, P.O. Box 30709, 00100, Nairobi, Kenya
3 International Potato Center, Sub-Saharan Africa Regional Office, ILRI Campus, P.O. Box 25171, 00603, Nairobi, Kenya
4 Jomo Kenyatta University of Agriculture and Technology, Department of Food Science, P.O. Box 62000, 00200, Nairobi, Kenya
5 Dedan Kimathi University of Technology, Private Bag 10143 Dedan Kimathi, Nyeri, Kenya
6 Kenya Excellence in Breeding Platform, CIMMYT, ICRAF Campus, P.O. Box 1041-00621, Nairobi, Kenya
7 Department of Food, Bioprocessing and Nutrition Sciences, North Carolina State University, Campus Box 7624 Raleigh, NC, 27695, USA

Firmness is a key quality trait influencing consumer preference of boiled sweetpotato roots. Identification of markers for firmness could enable breeders to effectively develop and screen for desirable clones in the breeding pipelines. To identify candidate genes for firmness, freshly harvested, mature, raw storage root transcriptomes of two hard, slow cooking varieties, NASPOT11 (cream-fleshed) and SPK004 (orange-fleshed), and two soft, fast cooking varieties, Wagabolige (cream-fleshed) and Resisto (orange-fleshed), were sequenced on an Illumina HiSeq2500 platform and analysed. In all the samples, more than 80 % of the reads mapped uniquely to the I. batatas haplotype resolved genome. There were 45 855 genes with detectable levels of expression. Of these, 7 635 were differentially expressed (padj < 0.05), with 3 633 significantly upregulated while 4133 were significantly downregulated in the hard, slow cooking varieties. Amongst the most significantly upregulated genes in the hard varieties were cell wall related genes; g62109; a pectin acetylesterase family protein, g35627; xyloglucan endotransglucosylase/hydrolase, and g4017; a cell wall-associated kinase. This strongly suggests that cell wall composition is closely associated with the development of firmness in boiled sweetpotato roots. This data provides a resource for genetic and genomic studies towards understanding and developing markers for the texture of sweetpotato roots.

Genomic analysis of liver abscesses in feedlot beef cattle

Melanie Hess[1], Galen Erickson[1] and Matthew Spangler[1]

1 University of Nebraska-Lincoln, Department of Animal Science, Lincoln, Nebraska, USA

Condemned livers are deemed not fit for human consumption, primarily due to the presence of abscesses, and result in over $15M in lost revenue to the US beef industry annually. The severity of liver abscesses is typically measured visually, using a 4-category system: 0 (no abscesses), A- (mild: 1-2 small abscesses), A (moderate: 2-4 active abscesses) and A+ (severe: 1 or more large, active abscesses). To investigate the underlying genetic control of liver abscess severity, we estimated genetic parameters and conducted a Genome-Wide Association Study for liver score using 1747 beef cattle fed a range of diets in a Nebraska feedlot. Diets were grouped based on corn type (Dry-Rolled Corn (DRC), High-Moisture Corn (HMC), DRC/HMC or Steam-Flaked Corn (SFC)) and byproduct (None, Modified Distiller Grains plus Solubles (MDGS), and sweet bran fed at 20%, 35% or 40%). Genotypes were available for all animals at 44,666 autosomal loci from the Illumina BovineSNP50 v2 chip that had a call rate ⩾ 0.9 and a minor allele frequency ⩾ 0.01. The heritability estimate of liver score was 0.10 ± 0.05. Within-diet heritability estimates were not significantly different from zero but ranged from 0.04 ± 0.04 (sweet bran fed at 40%) to 0.21 ± 0.21 (No byproduct). Results from a BayesB analysis with pi=0.95 suggest that liver score is a highly polygenic trait with no large QTL segregating in this population. The genetic and phenotypic correlations of liver score with hot carcass weight, 12th rib fat thickness, rib eye area, or marbling were not significantly different from zero but ranged from -0.17 ± 0.18 (12th rib fat thickness) to 0.22 ± 0.18 (rib eye area). Based on the heritability estimate reported herein, genetic selection for reduced liver score could be a useful tool to reduce the occurrence of liver abscesses in feedlot cattle, alongside current mitigation strategies such as carefully formulated diets. However, more data are needed to gain a better understanding of the genetic basis of risk of cattle getting liver abscesses, particularly the possibility of genotype x diet interactions.

Genomic biotechnologies as cattle breeding tools in Africa

Emily K. Kathambi[1, 2], Tad S. Sonstegard[2], Peter A. Larsen[1]

1 Department of Veterinary Biomedical Sciences, University of Minnesota
2 AcceligenTM, 1246 University Avenue West, #301, St. Paul, MN 55104, United States

The primary source of livelihood for approximately 75% of Africans is agriculture and agribusiness. Increasing livestock productivity in Africa requires simultaneous interventions in the areas of animal feed, health and genetics. African livestock breeds are diverse while the livestock systems are dynamic, with many small- and medium-scale systems. In 1999, it was noted that African countries stood to gain the most from modern biotechnology applications. Artificial insemination and selective bull mating have been aggressively used for dairy cattle breeding in Africa with impressive results. Adoption of these breeding technologies is dependent on; farmer's experience in conception success and management practices. In 2019 it was observed that genomic applications were benefiting African livestock systems in a variety of ways. Genomic selection using low-density SNP assays was used to estimate breed proportion and parentage assignment in crossbred dairy cattle in East and South Africa. Quantitative trait loci influencing response to trypanosome challenges were mapped in trypanosome tolerant West African N'dama cattle. Efforts are underway to discover genetic variants of economic significance in African Indigenous livestock populations. New genomic cattle technologies are being studied for introduction into the African region such as gene edited African cattle cell lines resistant to African Trypanosomiasis, bovine tuberculosis and tolerant to heat. Our intent is to evaluate the current state of cattle biotechnologies in Africa including new, existing and enhanced genomic technologies and their use as breeding tools. The review also aims to identify adoption rates of these biotechnologies and the reasons for the low uptake. We aim to evaluate the perceptions of the region towards gene edited cattle breeds, the social, cultural, political and economic implications these technologies compared to existing breeding technologies.

An automated, normalizing workflow for multiplexed next-generation sequencing (NGS) library construction in low-coverage whole-genome sequencing of crop specimens

Joseph Mellor[1], Charles Ariele Hanek[1], Marcel Brun[2], Jessica Smith[1], Matt Stull[2], Emma Schulze[2], Alex Watros[2], Joshua Hill[2], Mark Scheel[1], D. Johnson[2]

1 seqWell, Inc. Beverly, MA 01915 USA
2 Texas A&M AgriLife Genomics and Bioinformatics Service, College Station, TX 77845

Microarray-based genotyping has been the dominant method for genome-scale genotyping for over two decades. However, recent improvements in sequencing technology have made imputation of common genomic variants via low-coverage whole genome sequencing (lcWGS) a cost-effective alternative. This "low pass" approach to genome sequencing is particularly useful in agricultural population studies, as it allows for the cost-effective generation of genomic information on a large scale.

The use of lcWGS (long-read whole genome sequencing) as an alternative to array-based genotyping is cost-effective when performed on large-scale sequencing instruments that provide low cost per unit of data. However, this presents a challenge in generating a large number of lcWGS libraries to utilize the capacity of these large instruments. Developing highly multiplexed library construction methods is therefore important in order to fully take advantage of lcWGS for population-level genotyping of commercial herds or food crops.

To fully realize the potential of lcWGS in agricultural populations, it is important to develop highly multiplexed library construction workflows that can be used on large-scale sequencing instruments. These workflows should be able to generate multiplexed NGS libraries from genomic DNA, and efficiently automate the process of normalizing NGS libraries from typical starting samples of plant or animal DNA.

In this study, we evaluate an optimized automation workflow using seqWell's purePlex library preparation technology on a Sciclone G3 NGSx Workstation liquid handling workstation from PerkinElmer, Inc. We demonstrate the ability to generate highly multiplexed pools of lcWGS libraries from commercial samples of soy (Glycine max) for sequencing on an Illumina NovaSeq instrument.

Genomes to Genotyping

Large genotyping arrays for the identification of validated single nucleotide polymorphism (SNP) markers in major vegetable species

Martin W. Ganal[1], Joerg Plieske[1], Andreas Polley[1], Dagmar Kulosa[1], Thomas Gross[1], Eva Grafarend-Belau[1], Heike Gnad[1]

1 SGS Institut Fresenius GmbH TraitGenetics Section (SGS IF TG), Am Schwabeplan 1b, 06466 Seeland OT Gatersleben, Germany

The availability of high-quality SNP (single nucleotide polymorphism) markers is essential for the identification of suitable markers for mendelian and polygenic traits of agronomic importance that can be used for marker-assisted selection, the analysis of genome-wide polymorphism (e.g. for genetic relationship analysis), marker-trait association studies, and increasingly for genomic selection to improve complex traits such as yield. As source of SNP markers for major vegetable species, we have used publicly available genome sequences or we have generated novel genome sequences or transcriptome sequences (for some species). To identify and validate large numbers of functional and informative SNP markers for vegetable species, we have assembled several multispecies genotyping arrays on the Axiom platform that include these species: tomato, melon, watermelon, cucumber, lettuce, asparagus, onion and carrot. For each species approximately 10,000 to 75,000 high quality markers have been identified. By physical and, in part, genetic mapping, we have located these functional markers on the respective reference genome sequences. Examples will be presented for the use of these marker arrays in some species (e.g. carrot and asparagus) in order to identify markers linked to specific traits or for the analysis of genetic resources and breeding material. Currently, we are using these validated markers for the assembly of sets of highly informative SNP markers that can be used for the generation of smaller arrays in marker-assisted selection and genomic selection procedures at low costs in vegetable breeding.

Ultra-low-pass sequencing on the AVITI platform

Jesse Hoff

Gencove

The use of genomic prediction in agriculture has been shown to be an effective way to accelerate breeding by affordably providing accurate predictions of genetic merit. Genome wide genotypes of sufficient density and accuracy can substitute for phenotypic data collection in far less time. Several sequencing methods have been adopted to provide genotypes for these applications. Here we demonstrate an approach to ultra low-pass (ULP) sequence that will optimize cost, turnaround time flexibility and accuracy in a benchtop format. With a few dollars of sequence data it is possible to drive genomic improvement.

We aimed to evaluate the effectiveness of using a high throughput, ultra low coverage genotyping strategy using efficient library preparation, benchtop sequencers and low-pass sequencing combined with imputation. In total, libraries from 1,536 cattle samples were prepared using seqWell's plexWellTM Low Pass 384 kit and sequenced using the AVITI sequencer on two flow cells in a single run. We targeted average output of less than .2x coverage of the cattle genome. Data was demultiplexed and run on the Gencove analysis platform, resulting in imputation of over 60 million SNPs and indels. The genotype accuracy was calculated. We estimated the genomic relationship matrix of the ULP samples and compared to higher coverage data. We also evaluated the impact of further reductions in coverage, representing higher index counts.

The relative loss in genotyping accuracy from ULP would be low when compared to cost and throughput savings. A small lab with plant or animal samples can process up to 6,000 samples a week using this method, or more with additional indices. Our study demonstrates that the use of the Element AVITI system for genotyping in cows is a reliable and efficient method.

Dual-barcode feasibility in the AgriSeq™ targeted genotyping-by-sequencing workflow

Jennifer Lavoie[1], Angela Burrell[1]
Thermo Fisher Scientific

Targeted Genotyping by Sequencing (GBS) is a robust and cost-effective method for marker-assisted breeding and selection. Targeted GBS provides a scalable workflow, allowing for thousands of markers to be analyzed in up to thousands of samples per day.

The Applied Biosystems™ AgriSeq™ targeted GBS solution allows for up to 1536 uniquely barcoded samples to be processed in a single sequencing run. The current workflow utilizes a single-barcode approach, with 1536 barcodes available in 16 96-well plates. With amplicon coverage of 100X and using an Ion 540™ chip averaging 70M reads, up to 500 markers can be analyzed in 1536 samples on a single chip. Using an Ion 550™ chip averaging 115M reads, up to 750 markers can be analyzed in 1536 samples on a single chip. As GBS use expands, demand increases for higher throughput, faster turnaround, and increased multiplexing capability in GBS workflows.

Increasing available barcodes could increase the sample throughput of the AgriSeq™ workflow. In the current single barcode workflow, increasing available barcodes means increasing the number of 96-well plates to manage, which is undesirable. As an alternate solution for increasing throughput and barcode availability, we are exploring the use of dual barcodes. With this method, each sample would receive a combination of barcodes, allowing for more samples to be uniquely barcoded, which increases multiplexing capability and cost effectiveness without adding 96-well barcode plates. Here we report the feasibility of implementing the use of dual barcodes in the AgriSeq™ workflow and its implications for marker performance and coverage requirements.

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

Enhancement Of Oil Palm Sustainability With SHELL-Allele Supply Chain Screening

Chan Pek-Lan[1], A Rajinder Singh[1], Nur Nadia Kamil[2], Leslie Cheng-Li Ooi[1], Eng-Ti Leslie Low[1], Meilina Ong Abdullah[1], Robert A Martienssen[3], Ravigadevi Sambanthamurthi[1], Nathan Lakey[4], Mohamad Arif Abd Manaf[1] and Azman Ismail[2]

1 Malaysian Palm Oil Board, Advanced Biotechnology and Breeding Centre, No. 6, Persiaran Institusi, Bandar Baru Bangi, 43000 Kajang, Selangor Darul Ehsan, Malaysia.
2 Malaysian Palm Oil Board, Economics and Industry Development Division, Wisma Sawit, Lot 6, SS6, Jalan Perbandaran, 47301 Kelana Jaya, Selangor Darul Ehsan, Malaysia.
3 Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, 11724.
4 Orion Genomics, Saint Louis, Missouri, USA 63108.

The oil palm (Elaeis guineensis) is the world's most productive oil crop supplying 36.5% of the world's vegetable oil while only occupying 5% of land used to cultivate vegetable oil. A key oil palm domestication trait was a decreased thickness of the shell surrounding the kernel resulting in an increase in the fruit's oil-rich mesocarp layer. Oil palm's three fruit forms, low-yielding dura (thick-shelled) and pisifera (shell-less and female sterile), and high-yielding tenera (thin shelled), are controlled by any one of nine known mutant alleles of SHELL. The development of genetic maps, generation of a high-quality reference genome of Elaeis guineensis, and the short read sequencing of several extensive pedigrees segregating for the fruit form trait, analysed with homozygosity mapping, enabled the identification of SHELL, a type II MADS-box gene homolog of Arabidopsis SEEDSTICK (STK). The discovery has led to the innovation in SHELL-allele screening technology by identifying low-yielding contaminant (non-tenera) planting materials at the seed or nursery stage. A high throughput SHELL-allele assay was utilized to predict fruit form in nursery leaf and ungerminated seeds samples with near 100% accuracy. A total of 1,140,603 seed (33.7%) and leaf (66.3%) samples representing all oil palm growing regions and replanting supply chain segments across Malaysia were successfully tested. The results showed that only 87.2% of the tested materials are the high-yielding hybrid tenera form, while 12.8% comprised of low-yielding contaminant forms (dura or pisifera), which significantly exceeds the 5% maximum allowed by the national standard. SHELL-allele supply chain testing prior to field planting enables the removal of low-performing palm forms and increases the production of fresh fruit bunches by 5.6% and crude palm oil by 7.5% without the expansion of planted area. SHELL allele screening increases smallholder income by 10% nationally and adds $1.20BN USD annually in economic gains in Malaysia by identifying and removing the undesirable palms prior to field re-planting. We present for the first time, a new paradigm of 'screen-then-plant' cultivation of production materials, which leverages molecular precision agriculture to increase food production on existing plantations, improve sustainability and increase income of plantations and millions of subsistence farmers.

Bridging the Gap between Variant Genotyping and Scientists for Accelerated Innovations in Agriculture

Rajesh Perianayagam, VinothRaj Sekar, Ramesh Dharmaraj and Bill Perkins

Karyosoft Inc.

In simple terms, molecular genetics is the correlation of genotype and phenotype to discover the agronomically important genomic regions. With the direct integration of Next Generation Sequencing (NGS) technologies into in-house workflow, genotyping has improved dramatically in terms of the number of genome wide variants such as single nucleotide polymorphic (SNP) and Indel markers and thus the amount of genomic information. However, the current variant calling method is time consuming, labor intensive, requirement of computational skills, no suitable mechanism for managing millions and millions of SNPs and Indels in-house and more that delay innovation. At Karyosoft, we developed a cloud-based user-friendly platform Variants to circumvent these issues. Using this platform, we have reduced the variant calling time (end to end) from more than 28 hours per sample with 30x data coverage to less than 4 hours/sample. By incorporating parallel processing, we have reduced the time for multiple samples throughput further. Our Variants platform is 7 – 12 times faster, can reduce the cost by 12x – 7x and can save up to 168 days for 96 samples. Additionally, we have developed a cloud based highly scalable user-friendly Variant Mining Studio to manage and mine millions and millions of SNPs and Indels in seconds. Our platforms have a vast use in mutation discovery, direct genotyping, custom chip designing and amplicon sequencing. Above all, our user-friendly platform makes variant genotyping easy for scientists with any level of computational skills and empowers them to drive the innovations faster.

Innovations & New Applications

Investigation of genomic regions of interest in animal genomes using ONT adaptive sampling

Rebecca M Clarke[1], Andrew Hess[2], Alex Caulton[1], Rudiger Brauning[1], Kathryn McRae[1], Augustine Chen[3], & Shannon M Clarke[1]

1 Agresearch, Invermay, Mosgiel, NZ
2 University of Nevada, Reno, USA
3 University of Otago, Dunedin, NZ.

The challenge to provide energy-rich protein to the world while reducing the environmental impact is one of the largest challenges facing the animal industry today. To face this challenge, novel tools need to be adopted for methods to identify animals that perform optimally in their environment. This includes, but is not limited to, utilizing more complex variation, epigenetics, and microbes present within the host. A major hurdle in utilizing these data lies in the development of tools that can seamlessly capture multiple types of information in a high-throughput and cost-effective manner that can be implemented in industry. We propose that the sequencing platform developed by Oxford Nanopore Technologies (ONT) as a potential solution well-suited for this purpose.

In particular adaptive sampling, a software-controlled enrichment unique to nanopore sequencing platform, enables targeted sequencing of specific regions of a genome or species of interest, at higher coverage. This enrichment approach by ONT circumvents the need for upfront sample manipulation but also enables simultaneous capture of multiple sources of information such as methylation, mutations, and structural variances, in a single run. Using this ONT approach allows the capture of multiple sources of information in a single run with the aim of reducing the costs.
The inability to accurately impute complex variation is common across the genome and the direct capture of CNVs could provide improvement over the use of SNPs. The combination of long read technology and adaptive sampling could help target appropriate regions. We have utilized adaptive sampling to investigate regions of interest surrounding GWAS peaks and known copy number variation (CNV) regions in sheep. Sample multiplexing has been trialed to evaluate the possible use of adaptive sampling as a high-throughput cost-effective tool for the animal industry. This presentation will discuss the potential benefits and challenges currently being explored to integrate this tool into the animal industry.

Paragon AgriType Auto®: a targeted genotyping-by-sequencing (GBS) solution compatible with automation for high-throughput scalability

Jack Dong[1], Julia Spencer[1], Lucie Lee[1], Fang Xie[1], Michael Clark[1], Kalyani Patankar[1], Vidushi Kapoor[1], Lily Liu[1], Julia Yoo[1], Jeffrey Liu[1], Yutao Fu[1], Zhitong Liu[1], David Debruyne[1]

1. Paragon Genomics, Inc., Fremont, CA 94538

Next Generation Sequencing is becoming instrumental to advance agrigenomics research, where genome characterization proves crucial to optimize and render crop production systems more productive and sustainable in order to meet our ever-growing global food demand. Indeed, genotyping innovations drove further development of molecular breeding strategies, with key applications such as trait mapping, marker-assisted selection, or genetic engineering. However, high-throughput and cost-effective methods are required to maximize genotyping effectiveness in up to thousands of samples per day, as we face modern challenges such as population expansion and climate change. Targeted solutions represent ideal candidates, by enabling sample multiplexing at an unprecedented scale while remaining highly informative. However, current workflows are still hindered by time-consuming steps and limited automation compatibility to fully empower large-scale projects.

Here, we present a novel genotyping by sequencing solution, AgriType Auto®, a fast, streamlined and automation-compatible library preparation workflow, with high-throughput capability (96- / 384-well plates). In summary, we adapted our targeted, multiplex PCR technology to reduce turnaround time and increase speed as well as ease of use, allowing for higher throughput and adapted to automated handling. This novel method was validated in several distinct crops, using panels targeting up to 7,000 markers, highlighting leading-edge performance, with sequencing uniformity and call rate accuracy comparable to our CleanPlex technology®. These results hereby confirm the successful streamlining of our NGS technology to effectively support large-scale agrigenomics studies using molecular breeding to bring about crop improvement.

Scalable sample extraction and multiplexed library preparation for high-throughput crop genotyping via low-coverage whole-genome sequencing

Ariele Hanek[1], Michelle Rahardja[1], Jessica Smith[1], Joseph Mellor[1], Tobias Schughart[2], Maximilian Weiter[2], Jesse Hoff[3]

1 seqWell Inc., 66 Cherry Hill Dr. Beverly, MA 01915
2 BioEcho Life Sciences, BioCampus Cologne, Nattermannallee 1. 50829 Köln Germany
3 Gencove 30-02 48th Ave. Suite 370, Queens, NY 11101

The reduced cost and improved scale of DNA sequencing has permitted a variety of genotyping assays to be approached with sequencing-based alternatives over legacy approaches such as microarray-based genotyping. One key remaining bottleneck to unlock the efficient use of sequencing instruments for high-throughput genotyping is the creation of scalable workflows and chemistry for sample extraction and library prep.

In this study, we evaluate the use of the EchoLution platform (BioEcho Life Sciences) and plexWellTM Low Pass multiplexed NGS technology (seqWell) for performance in a routine crop genotyping context. A collection of 48 samples from tomato, maize and soybean seeds (inclusive of biological and technical replicates) was extracted to produce purified gDNA, which was then prepared as an Illumina-compatible multiplexed NGS sequencing library for sequencing on a NextSeq2000 sequencing instrument. The multiplexed library was sequenced to an average depth of 1x to facilitate low-pass imputation.

The resulting sequencing data was characterized for quality and uniformity between samples and by Gencove's imputation pipeline. The results of this study demonstrate the cost-effective feasibility to generate high-quality genotyping information from plant specimens such as tomato at large multiplexed scale.

Challenges of high-throughput phenotyping in a rangeland environment

Andrew Hess[1], Scott Huber[1], John Bergeron[1], Gary McCuin[1], Melanie Hess[2], Tracy Shane[1], Jason Karl[3], Mike Cox[4], Robert Washington-Allen[1]

1 University of Nevada, Reno, Department of Agriculture, Veterinary, and Rangeland Science, Reno, NV, USA
2 University of Nebraska-Lincoln, Department of Animal Science, Lincoln, NE, USA
3 University of Idaho, Department of Forestry, Rangeland, and Fire Sciences, Moscow, ID, USA
4 Nevada Department of Wildlife, Reno, NV, USA

Extensively managed livestock spend a significant amount of the year grazing a resource-limited environment where they are exposed to a multitude of environmental stressors, including extreme weather, predators, and varied terrain. Capturing the ability of an individual to cope with these stressors has the potential to improve animal welfare and productivity in these environments. To do this, measurements of resilience and efficiency will be a useful tool to make continued improvement in rangeland livestock through genetic selection. These measurements need to occur over the course of the animal's time in such an environment or capture the sum of the animal's ability to cope with the environment. The ability to capture such data is encumbered by limited-to-no access to commodities such as power and internet, and limited manpower for manual data collection. Therefore, such tools need to be self-sustaining, automated, and cost-effective to be able to capture data on a large number of individuals. To this end, we have been exploring the use of precision livestock tools to capture sheep performance in a rangeland environment. We are currently working on developing a portable walk-over-weighing system to automatically capture an individual's weight that uses a combination of solar powered portable batteries and an RFID system with the aim to make a compact weighing system that can easily be transported in rough mountainous terrain. We are also using low-cost commercial-off-the-shelf GPS units to capture animal behavior in these environments. Through these data, we aim to develop traits that integrate animal performance and behavior to identify animals that are better suited to a rangeland environment. We expect that breeding of such animals will result in a positive impact on productivity, a reduced impact on the ecosystem, and lower mortality rates, thereby improving the sustainability of extensive livestock production systems.

A transformational one-step library preparation method for multiplexed plasmid and amplicon sequencing

Curtis Knox[1], Jack T. Leonard[1], Rebecca Feeley[1], Stella Huang[1], and Jessica Smith[1]

1 seqWell Inc., 66 Cherry Hill Dr. Beverly, MA 01915

Conventional library preparation methods demand multiple incubations interrupted by cumbersome pipetting steps. Here we describe a transformational new library prep technology enabling barcoding and amplification of 8 - 384 dual-indexed libraries in 90 minutes in 16 µl "one-pot" reactions, while simultaneously auto-normalizing library read-count and insert size. The simplicity of the ExpressPlex™ Library Prep workflow makes it uniquely well-suited for manual, automated and ultra-high throughput (UHT) library prep from synthetic constructs such as plasmids and PCR products. Read-count normalization data is easily achieved without individual sample quantification. To demonstrate miniaturization, we prepared plasmid libraries on the SPT LabTech mosquito® HV at 3,200 nanoliters per reaction in 384-well PCR plates, and the sequencing data were used for de novo, full plasmid assembly. This revolutionary library prep method will enable more efficient discovery in all aspects of genomics, including synthetic biology, plant, animal, and microbial research.

Plant inter-cellular and inter-species single-cell omics

Marc Libault

1 Department of Agronomy and Horticulture
2 Center for Plant Science Innovation,
3 Single Cell Genomics Core Facility
4 Center for Biotechnology
5 University of Nebraska-Lincoln

Each plant cell is different based on its biological function, developmental stage, and interaction with its environment. The unique biology of each plant is reflected in the differential use of the genomic information and activity of the genes. Plant single-cell-omic technologies recently emerged to capture cell and cell-type-specific transcriptomes and profiles of chromatin accessibility. In our lab, we have implemented different strategies to capture and interpret single-cell molecular features. Our research supports the long-term objective of the Plant Cell Atlas community (https://www.plantcellatlas.org): access, interpret and understand molecular datasets at the single-cell level.

Here, we will present our work on isolated plant nuclei from different plant species and organs, the better understand the activity and role of plant genes. Specifically, we will present the current status of various single-cell technologies [i.e., single-nucleus RNA sequencing (sNucRNA-seq), single-nucleus Assay for Transposase Accessible Chromatin sequencing (sNucATAC-seq), single nucleus multiomic technologies, and spatial transcriptomics] and provide examples of different biological outcomes of our work. For instance, we will describe our current work establishing the single-cell resolution transcriptome atlas of the soybean plant and the use of Molecular Cartography and laser-capture microdissection technologies to support the annotation of the soybean cell types according to their transcriptomic profiles. We expect that this resource will support functional genomic assays and the development of synthetic biology strategies, We will also present integrated analyses between single-cell transcriptomic and metabolomic/proteomic datasets to reveal the differential activity of metabolomic pathways and proteomic features between plant cell types. Ultimately, this new fundamental knowledge will enhance our understanding of gene function, gene networks, metabolomics pathways, and traits to enhance crop biology.

Genome-edited organisms: options for traceability

Hellen Mbaya[1, 2, 3], Alan Raybould[1], Simon Lillico[2, 3], Steve Kemp[3, 4], Geoff Simm[1]

1 Global Academy of Agriculture and Food Security, the Royal (Dick) School of Veterinary Studies, UK
2 The Roslin Institute, University of Edinburgh, UK
3 Center for Tropical Genetics Livestock and Health, Edinburgh, UK
4 International Livestock Research Institute, Nairobi, Kenya

The advent of genome editing (GEd) has provided new and more efficient ways to introduce targeted changes in the genomes of both plants and animals. This has resulted in use of the technology by a wider variety of stakeholders for various applications in comparison to genetic modification (GM) technology. Globally, regulators in different parts of the world are now examining their current frameworks to assess their applicability to this new technology and products. Techniques currently used for detection of inserted transgenes in GM products include DNA and protein-based methods, which are applied based on a country's financial capability and infrastructure availability. Products made from GEd crops and animals are now getting approval for commercialisation with the most recent being GEd tomato, sea bream and tiger puffer fish in Japan. Questions remain regarding whether GEd products will be subjected to the same traceability requirements as GM products or whether new ones will be put in place. For this study, sgRNAs targeting the prolactin gene (PRLR) region in the sheep genome were designed. Further, a restriction site for EcoRI enzyme (DNA footprint) was incorporated next to our target within the genome during design of homology directed repair (HDR) templates. Thereafter, two GEd transfection techniques Ribonucleoprotein (RNP) and plasmid transfection were applied to introduce the DNA footprint to the genome targeting the area around our gene of interest. Finally, PCR-RFLP and PCR-sanger sequencing were used for detection and validation of GEd events in ovine embryonic fibroblast cells. Results showed four clones potentially homozygous for the HDR events and four were wild type. Our study shows some of the techniques that regulators can apply in the traceability of GEd organisms in both high resource and low resource settings.

Pick your poison: benchmark comparison of sample storage buffers (ethanol and RNAlater), time (0-6 weeks), and temperature (22°C and 4°C) for long-read sequencing of animal and plant genomes

Jeremiah J Minich[1] *, Malia Moore[1] *, Anthony Alyward[1], Nick Allsing[1], Emily Murray[1], Loi Tran[1], Nolan Hartwick[1], and TP Michael[1]

1 The Molecular and Cellular Biology Laboratory, The Salk Institute for Biological Studies, La Jolla, CA, 92037, USA

There are over 1.6 million species of animals and 435 thousand species of plants on Earth. Long read sequencing technology is enabling biologists to digitally archive the genomes of these organisms to understand fundamental basic questions in evolution, bolster conservation strategies in the midst of massive extinctions, and improve food production through crop and livestock genomics. One of the primary challenges however, is obtaining samples from field sites and preserving them in a way that ensures high quality DNA for long read sequencing. Here we demonstrate the impacts of storage time (0 days, 1 week, 3 weeks, and 6 weeks), storage temperature (4 oC and 22 oC), and storage buffers (95% EtOH and RNAlater) on the preservation of fish blood (n=10 species) and plant tissue (n=5 species). We compare how storage methods impact DNA QC metrics such as total yield, purity (260/280 and 260/230), and fragment size (DNA conc. CV, femto pulse average bp, and femto pulse fragments above 50 kb). In addition, we compare how storage methods directly impact sequencing performance on 131 libraries (fish=90, plant=41) using the early access ONT LSK114 high accuracy chemistry and evaluate performance metrics such as read N50, read mean quality, and total sequencing yield. Finally, we evaluate to what extent DNA QC metrics predict sequencing performance. To validate our findings, we sequenced (ONT LSK114 with Super Accurate Basecalling) and assembled genomes from three fish species at >30x coverage with samples stored in 95% EtOH at 22 oC for 6 weeks and one plant species at >30x coverage stored in 95% EtOH at 4 oC and RNAlater at 22 oC for 3 weeks.

Blossoming Dutch technologies: application of graph-based pangenomes and high-quality genome assemblies to accelerate plant breeding

Nijland Bart, **van Dam Peter, van Kruistum Henri, Brenes Guallar Megan, de Heer Mark, van de Rhee Miranda, van Belzen Almer, van Bers Nikkie, Ferreira Stephan, Berke Lidija**

Novel developments in long read sequencing and bioinformatics accelerate the generation of high-quality reference genomes, which provide an essential basis for molecular plant breeding. The availability of multiple high-quality reference genomes per species is becoming cost-effective and allows for the capture of all variation in a crop species using a pangenome approach. Genetwister has a wide experience in developing high-quality genome assemblies of a diverse array of plant species; from small diploid genomes to large and highly complex polyploid genomes. In this presentation, the application of high-quality genome assemblies to advance molecular plant breeding will be highlighted using two example cases.

The tulip is a symbol of Dutch culture and has high commercial value all across the world. We have tackled the de novo assembly of its complex genome, which is extremely large in size (35 Gb) and has more than 80% of repeat content. Our assembly is of superior quality to other publicly available genome assemblies of similar size and complexity.

The second showcase describes the construction of a graph-based pangenome for improved identification of different types of variants. Using the PacBio HiFi long-read sequencing technology, we generated high quality reference genomes of different types of cucumber varieties. The resulting assemblies are highly contiguous, all consisting of full-length chromosomes or chromosome arms. We integrated these assemblies into a graph pangenome which allowed the mining for novel genetic variation to use in accelerated variety development. These variants ranged from large structural variants and presence/absence polymorphisms, to small indels and SNPs. In our pangenome approach, additional varieties can easily be added and compared by mapping resequencing data to the pangenome graph and calling variants. Integration of multiple data types (variants, annotation and phenotype information) using a pangenome approach is the first to resolving complex traits for plant breeding.

Development of low-density single nucleotide polymorphism (SNP) genotype panels for parentage verification in South African Bonsmara and Drakensberger cattle

Yandisiwe P. Sanarana[1,2], Donagh P. Berry[1,3], Azwihangwisi Maiwashe[2], Cuthbert Banga[2], Este van Marle-Köster[1]

1 University of Pretoria, Department of Animal and Wildlife Science, Pretoria, Hatfield, 0002, South Africa
2 Agricultural Research Council-Animal Production, Pretoria, Irene, 0062, South Africa
3 Teagasc, Animal and Grassland Research, and Innovation Center, Fermoy, Cork, P61 C996, Ireland

Genotyping panels exploiting single nucleotide polymorphism (SNP) markers have superseded the use of microsatellites for parentage testing and other applications. Although high-density SNP arrays allow accurate parentage assignment, they are not practical for routine application in some settings. Constraints include high cost, processing time and analyzing the samples. In this study, low-density SNP panels with high resolution power were developed to perform accurate parentage tests for the South African Bonsmara (BON) and Drakensberger (DRB) cattle breeds. Genomic data consisting of 38 888 variants from the GeneSeek® Genomic Profiler (GGP) Bovine 150K BeadChip was available for 2 835 animals representing both breeds. The data included 161 sire-progeny and 313 dam-progeny pairs that were previously validated using 119 375 SNP markers. Informative SNPs were selected based on high minor allele frequencies (MAFs), clustering quality, and high call rates. To reduce linkage between markers, SNPs were spread across and within chromosomes with a minimum distance of one mega base pair (Mb) apart. A total of 200 markers were selected to develop a multi-breed and population-specific low-density genotype panels. On average, the genotype panel with SNPs selected across the breeds had a lower minor allele frequency (MAF) of 0.40 while breed-specific panels had higher MAF with an average of 0.48 and 0.49 in the DRB and BON, respectively. The panels were validated on parent-offspring pairs based on the method of parentage exclusion of sharing at least one allele between parent–offspring pairs. The results revealed that breed-specific SNP marker panels had high accuracy for parentage exclusion without any false-negatives, while the multi-breed panel did not perform well. There were no parental-offspring discrepancies in the DRB when the multi-breed panel was used, while 4.2% false-negatives were observed in the BON. This study demonstrated that markers selected based on multiple breeds may not be sufficient, and at times may fail to exclude parentages of inbred relatives. Furthermore, this study demonstrated the feasibility of developing low-density SNP panels, which could be utilized to enhance livestock productivity in developing countries where genetic improvement through selective breeding has been hampered by a general lack of pedigree data on selection candidates.

A visual approach to assist mating processes in livestock

Alana Selli[1], Steve Miller[2], Ricardo Vieira Ventura[1]

1 School of Veterinary Medicine and Animal Science of University of São Paulo, Department of Animal Nutrition and Production, Pirassununga, São Paulo, Brazil.
2 University of New England, Animal Genetics and Breeding Unit, Armidale, New South Wales, Australia.

In the last fifteen years, genomic selection has benefited from the development of low and medium density genotyping panels, a lower cost solution, allowing breeders to genotype larger numbers of animals. Traditionally, most mating tools make use of Estimated Breeding Values (EBV) in association with inbreeding coefficient thresholds. Additionally, data visualization can provide valuable insights, either through ready to use and paid software, or via Open Source options, which require coding skills. The objective of this study is to develop a visual pilot tool which aggregates information from haplotypes presenting significant marker effects, contrasting these locations among the top ranked individuals from the whole population, and a set of candidates (selecting one cow at a time and updating a list of sires ranked according to their EBVs and inbreeding information). This visual approach requires a set of phased genotypes from both groups of male and female candidates, and marker effects obtained typically from Genome Wide Association Study (GWAS) investigations. The phased genotypes are converted via Python language into fixed sized windows, which are further used to compare the similarity within windows between the mating candidates. Including marker effects allows to select only segments of higher importance for a given trait, and each sire candidate is also evaluated according to the incidence of this set of haplotypes in its genome profile. Our pilot also considers the definition of top ranked individuals to have their haplotypes used as base comparison, where all the candidate haplotypes will be contrasted against. The Dash and Plotly libraries were used to create a visual interface which allows the interaction between the user and the data. Outcomes from this approach include tracking haplotypes from close relatives, comparing haplotypes from a given individual to the rest of the population, and obtaining their incidence in the population. Further studies are necessary to assess the advantages and disadvantages of using the chosen libraries opposed to readily available Business Intelligence tools, such as Tableau or Power BI, when using big data. Improvements in visual components such as the addition of graphics and ranking options are necessary and currently under development.

Targeted stimulation of meiotic recombination for crop improvement using novel breeding technology

Gaganpreet Sidhu[1], Prasenjit Saha[1], Minghui Wang[1], Malika Sharma[1], Adam Doniger[1], Alexandre Serero[1], Alain Nicolas[1], Luc Mathis[1], Wojtek Pawlowski[1,2]

Rapid increase in world population poses major challenges including a global scarcity of food. Therefore, it is crucial to grow crops that are higher yielding, disease/pest resistant and are resilient in different climate conditions. Traditional crop improvement methods take many years and are oftentimes unsuccessful due to lack of recombination in the desired genomic regions. Thus, a technology targeting meiotic recombination in recombination-poor regions will be important. Meiotic recombination is a conserved mechanism among sexually reproducing organisms. It is initiated by the Spo11-TopVIB protein complex that induces a large number of programmed double-strand breaks (DSBs) that are repaired as crossovers (CO) or non-crossovers. Meiotic recombination plays a critical role in crop improvement as it creates novel gene combinations. Breeders apply selection on the diversity created by recombination to increase genetic gains. The formation of COs is non-random with most COs forming on terminal regions of chromosomes. Therefore, a plethora of genes in plants and their wild relatives are inaccessible as they are in recombination poor regions.

We developed a novel technology to target meiotic recombination, based on the expression of a synthetic dCas9-Spo11-1 protein fusion. Modeling this methodology in yeast, we were able to significantly increase CO numbers in several recombination poor regions. We are now applying this technology to agriculturally important crops. Since the distribution of recombination events is governed by several factors such as chromosomal location, DNA methylation and chromatin openness, we use Artificial Intelligence (AI) to dissect high-resolution datasets for recombination, and train the machine algorithm program to predict the best genomic sites for efficient targeting of fusion protein. We aim to use this technology to (i) create genetic diversity in the low recombination genomic regions, (ii) remove linkage drags and (iii) accelerate the introgression of new allele(s) from wild relatives into elite cultivars.

Knocking out of banana U-box polyubiquitin ligase genes to enhance resistance to Xanthomonas wilt disease

Jaindra Nath Tripathi, Valentine Otang Ntui, Trushar Shah, Leena Tripathi

International Institute of Tropical Agriculture (IITA)

Banana Xanthomonas wilt (BXW), caused by Xanthomonas campestris pv. musacearum (Xcm), a deadly disease prevalent in African countries, wiping out banana (Musa spp.) cultivation. All the cultivated banana varieties are susceptible to BXW disease. CRISPR/Cas technology has the potential to provide durable solutions to improve crops by enhancing production. Our objective was to develop a genome-edited banana with resistance to BXW by editing Musa polyubiquitin box proteins (PUB) targeting genes. PUB22 and PUB23 genes encode a cytoplasmically localized U-box domain containing E3 ubiquitin ligase that acts as a negative regulator of pathogen associated molecular patterns (PAMP) triggered immunity in plants. Comparative transcriptomic data showed that PUB22 and PUB23 genes were suppressed in BXW-resistant wild-type banana genotype Musa balbisiana, whereas upregulated in BXW-susceptible banana cultivar in response to bacterial infection. Therefore, a multiplexed CRISPR/Cas9 plasmid targeting the PUB22/23 genes was designed and inserted into embryogenic cells of BXW-susceptible banana Sukali Ndiizi (AAB genome), and generated edited events, which were characterized for targeted mutations by sequence analysis. These events showed the targeted mutations with small indels and large deletions (Δ> 83 bp). These mutants showed enhanced resistance against Xcm upon artificial infection under greenhouse conditions. Knocking out the banana U-box polyubiquitin ligases is a potential strategy to develop disease resistance bananas. This technology can be used similarly to other crops to develop disease resistance varieties by knocking out through CRISPR-Cas9 technology.

Ultra-high throughput multi-omic analysis for agrigenomics on PacBio Revio

Jeffrey Burke[1], Renee Fedak[1], Victor Llaca[2], Charlotte Harris[2], Gina-Zastrow Hayes[2], and Kelvin Liu[1]

1 PacBio, Menlo Park, CA
2 Corteva Agriscience, Johnston, IA

Climate change and the rapidly growing global population is increasingly straining world food production. Long-read sequencing is currently being used in large-scale agricultural initiatives to help improve crop yields and combat pests and disease to meet increased agricultural demand. This is being driven by recent technological advances resulting in increased throughput and reduced sequencing costs.

We present an ultra-high throughput, sample-to-answer solution for agrigenomics built around the Revio system from PacBio through a partnership with Corteva Agriscience. This solution includes fully automated HMW DNA extraction, shearing, library prep and size-selection for multiplexed sequencing, allowing users to produce thousands of plant genomes per year at roughly $10/Gb. The Revio system can generate 360 Gbs of HiFi data (90% bases ≥ Q30) per day, offering a complete view of the genome, epigenome, and transcriptome.

Utilization of this newly developed, ultra-high throughput workflow resulted in multiplexed sequence data from several maize lines on multiple Revio SMRT Cells. This data was assembled and compared to results previously generated on the Sequel IIe platform.

Other

The repeatome and satellitome landscapes of Calophyllum soulattri and C. inophyllum yield significant insights into speciation and possible chromosomal markers for intra- and interspecific variability of Calophyllum spp. Genomes

Other

POSTER 700

Charles Anthon E. Cadorna[1], Dexter G. Pahayo[1], and Jessica D. Rey[1]

1 Plant Molecular Phylogenetics Laboratory, Institute of Biology, University of the Philippines, Diliman, Quezon City, Metro Manila, Philippines 1101

Modern plant breeding research requires a large amount of data to function effectively. Data repositories are improving in their ability to store this data, but there is a growing need for interoperability between disparate data sources and applications. The Breeding Application Programming Interface (BrAPI) project offers a solution to this problem with a standardized RESTful web service API specification. This specification provides a standard data model for the plant breeding domain, plus a well-defined set of methods for interacting with the data. The goal of the project is to promote interoperability, data sharing, and open source code sharing across organizations who produce and consume data in this domain. The BrAPI project is a community built project and that community is well established and continuously growing. The standard is built based on concrete use cases to solve real interoperability challenges faced by the community. Beyond the core standard, the community has built a variety of open source tools and resources to help build and test implementations of the specification. The community is also constantly producing new BrAPI compliant applications, analysis tools, and visualizations that will work with any BrAPI compliant data source.

This presentation is intended for people in the overlap between plant scientists and research data management. It may be interesting to anyone interested in APIs, data sharing strategies, and collaborative software development.

CERCA: Circular Economy that Reimagines Corn Agriculture

Other

POSTER 701

M. Cinta Romay[1], CERCA team, Edward S. Buckler[1,2,3]

1 Institute for Genomic Diversity, Cornell University, Ithaca, NY 14853, USA
2 School of Integrative Plant Sciences, Plant Breeding and Genetics Section, Cornell University, Ithaca, New York, 14853, USA
3 USDA-ARS, Robert W. Holley Center for Agriculture and Health, Ithaca, NY 14853, USA

Maize is the number one production crop in the US and globally, supporting the global food supply as well as portions of the energy system. While maize is part of highly efficient agricultural systems, it contributes to climate change, especially through greenhouse gas emissions. Nitrogen fertilizer is the largest energy input for maize; moreover, excess nitrogen results in water pollution and contributes 6% of the entire US GHG (55% of agricultural emissions) through nitrous oxide release. Improvements to genetics and management practices that would minimize agriculture's contributions to climate change are necessary to ensure the sustainability of agricultural systems. To start solving this problem, a team of 27 scientists from across the US is working together. This research aims to create a Circular Economy that Reimagines Corn Agriculture (CERCA) - converting maize to an earlier season annual with reduced environmental impacts through increased uptake and recycling of nitrogen and phosphorus fertilizer. The team is organized in three highly integrated activities: (1) Modeling of plants, farms, environments, and economics to determine the most important combinations of traits and environments likely to benefit from new cropping systems. (2) Trait Discovery through evaluation of related wild species and maize landraces that recycle nutrients and have some cold tolerance. (3) Trait Development through stacking and testing of genetic improvements predicted to have substantial impact. This project will start with seven native trait concepts focused on greater on-farm nutrient recycling: Nitrogen remobilization to roots, reduction of Nitrogen and Phosphorus in the kernel, Biological Nitrification Inhibition (BNI), germination and seedling cold tolerance, and seedling establishment in cool temperatures. With these traits, CERCA reduces maize environmental impact by reimagining the cycle of planting and fertilization without changing most of the existing equipment and processes, thus making adoption more likely and efficient.

Optimized nuclei isolation from diverse angiosperm species

Joshua C. Wood[1], Chenxin Li[1], Natalie C. Deans[1], Anne Frances Jarrell[1,2], Dionne Martin[1,3], Kathrine Mailloux[1], Yi-Wen Wang[1], C. Robin Buell[1,2,3]

1 Center for Applied Genetic Technologies, University of Georgia, Athens, GA, USA
2 Institute for Plant Breeding, Genetics, & Genomics, University of Georgia, Athens, GA, USA
3 Department of Genetics, University of Georgia, Athens, GA, USA
4 Department of Crop & Soil Sciences, University of Georgia, Athens, GA, USA

Generating quality nuclei is key to the success of a multitude of genomics workflows including High Molecule Weight (HMW) DNA isolation, Hi-C, Assay for Transposase-Accessible Chromatin using sequencing (ATAC-seq), and single cell sequencing (e.g., single nuclei ATAC-seq and single nuclei RNAseq). However, the presence of the plant cell wall poses a challenge in the isolation of nuclei from all plant tissues. We have demonstrated a simple, quick, optimized protocol for isolating nuclei from diverse angiosperm species, which includes, maize, soybean, tomato, potato, and wheat. This protocol has been tested on multiple tissue types as well as fresh and frozen samples. High quality nuclei are obtained via chopping in an isolation buffer on ice, where a key factor affecting nuclei integrity is the concentration of detergent (Triton X-100). Nuclei purity is attained through a series of serial filtrations followed by fluorescence activated nuclei sorting to remove debris and plastids. This method uses relatively little tissue, ~300 mg, and yields at least 100,000 purified nuclei in as little as twenty minutes. We have observed that optimal detergent concentration varied across the species tested and sometimes within species when different tissues were used. We also observed no pattern when considering genome size or taxonomic group, highlighting the importance of per species and tissue sample optimization. Our protocol and guidelines provide a valuable resource for optimizing nuclei isolations from novel plant species and tissues.

Phenotyping Biometrics Where the Rubber Meets the Road

Development of a semi-automated method for phenotyping awns in Miscanthus

Wren Jenkins[1,2], Pradeep Hirannaiah[1], Xianyan Kuang[3], Ernst Cebert[3], Brian S. Baldwin[4], Jesse Morrison[4], Erik J. Sacks[5], Therese Mitros[6], Daniel Roksar[6], Shengqiang Shu[7], Kerrie Barry[7], Chris Plott[1], Jerry Jenkins[1], Jeremy Shmutz[1], Jane Grimwood[1], PacBio, Kankshita Swaminathan[1]

1 HudsonAlpha Institute for Biotechnology, Huntsville, Alabama, United States of America
2 University of Alabama, Huntsville, Alabama, United States of America
3 Alabama A&M University, Huntsville, Alabama, United States of America
4 Mississippi State University, Mississippi State, Mississippi, United States of America
5 University of Illinois Urbana-Champaign, Champaign, Illinois, United States of America,
6 University of California, Berkeley, California, United States of America

An awn is a stiff bristle-like structure found in the flowers of many grass species in the family Poaceae. Awns are involved in self-burial of the seed, and have been found to have correlations to other agronomic traits of interest. This study pursues phenotyping and genetic mapping of awn characteristics in Miscanthus, a C4 perennial grass of interest due to its high biomass yields over consecutive growing seasons and ability to survive on more marginal land. Among Miscanthus species, the presence of an awn is a M. sinensis trait. M. sacchariflorus lines lack awns. An F1 and F2 population was developed using M. sinensis (PMS-014) and M. sacchariflorus ssp. Lutarioriparius (PF30022), and planted in a replicated field trial (n=4) in two locations, Alabama A&M, Alabama, and MSU, Mississippi. Due to their small size, Miscanthus awns are challenging to phenotype. To improve this, a method was developed for automated length measurement via a python image analysis code. Input images for this program needed to present high contrast between the desired objects-to-measure and their background, as a binarizing step was performed to generate a black-and-white image. To ensure the greatest contrast, microscope slides were painted to produce a matte black background, allowing sharp distinction between the awns and their background. A microscope with an attached camera was used for imaging and application of a reference scale bar. After binarizing, the image was skeletonized to reduce the awn structures and the scale bar to singular lines of pixels, which were then able to be measured using an arc length function to identify awn size. In order to identify measurements of specific awns, and compare them relative to the scale bar to provide standardized measurements, image contouring was used to isolate specific image elements. The outputs of this code were compiled and mapped against the Miscanthus genome in order to create a Quantitative Trait Locus map of awn-affecting genes. This found that chromosomes 3 and 6 of Miscanthus carry genes having measurable effect on awn length.

Development of genomic evaluation of a breeding value system for milk production traits with on-laboratory quality control of phenotypic data

Rukin I.V.[1], Krutkina M.S., Pantiukh K.S.[1], Kamaldinov E.V. [2]

1 JSC «Agroplem», Moscow, Russian Federation
2 Novosibirsk State Agrarian University, Novosibirsk, Russian Federation

Practical experience with implementation of dairy cattle breeding systems has shown that the value of breeding progress depends on reliability of phenotypic data. Among the most dairy selection systems in different countries, consortium of dairy industry organizations plays significant role in exclusion of primary phenotypic data reliability bias. For dairy traits, primary phenotypic data are obtained on the dairy farm and in specialized milk analysis laboratories. The data are then sent through associations to data calculation centers, where it is already used to calculate genetic and genomic breeding values. Quality control of all processes controlled through the introduction of the DHI system. However, the DHI system was not created and implemented in the Russian Federation, which does not guarantee the data reliability. In this work we have developed an evaluation the breeding value system of dairy cattle fully processing in one structural organization, the laboratory cluster Agroplem: from control milking to calculating the genomic evaluation of breeding value based on it. For two years, 1,832,278 test day milk samples from 420,400 Holstein breed cows from 95 farms were tested. All the TD data was filtered through the 4 steps: (1) estimation of falling within the confidence interval, (2) "spill out of the tank" estimation, (3) filtering of sample with signs of souring, and (4) filtering of samples with signs of water dilution. Based on the filtered data and on 7745 female genotypes, the estimation of breeding value for 5 milk production traits was calculated using ssGBLUP. The developed evaluation of breeding value system made it possible to evaluate both imported and domestic bulls based on reliable data of their progeny. Evaluation of breeding value was obtained for 25,660 bulls, among which 610 bulls had genetic evaluation with reliability higher than 70%. This level of reliability allows us to use this system in selection of bulls for AI stations. The reliability of genomic evaluation of breeding value averaged 45%. This indicator is insufficient for use in routine breeding selection programs. To increase the reliability, it is planned to further expand the reference population by genotyping animals with the highest reliability level.

Genetic regulation of self-organizing canopy patterns and their impacts on light interception in maize

Yan Zhou, **Aaron Kusmec, Patrick S. Schnable**

Iowa State University

The efficiency of canopy interception of solar radiation is a major contributor to the photosynthetic efficiency of crop plants. Light interception is itself a function of canopy architecture, including, leaf number, length, width, and angle, as well as azimuthal canopy orientation(s). We report on the ability of some maize genotypes to alter the orientations of their leaves during development in coordination with adjacent plants. Although these genotypes generally retain the typical alternate-distichous phyllotaxy of maize, their leaves grow parallel to those of adjacent plants of the same genotype. Previous morphological studies suggest this "parallel pattern" phenotype may be a response to the ratio of red to far-red light (R/FR), and hence a component of the shade avoidance syndrome (SAS). A genome-wide association study (GWAS) conducted on the parallel pattern phenotype identified candidate genes, many of which have been reported to be associated with SAS, including phytochromeC2 (phyC2). In addition, GWAS on the fraction of photosynthetically active radiation (PAR) intercepted by canopies (designated "PAR interception") of the diversity panel identified genes known to regulate leaf development, including liguleless1 (lg1). Mutants in both the phyC2 and liguleless1 genes exhibit altered canopy patterns, such that the numbers of interrow leaves are greatly reduced as compared to non-mutant controls. This results in dramatically decreased PAR interception by the lg1, lg2 and Lg3 mutant canopies, demonstrating a novel function of these genes, which were originally defined by their ability to regulate ligule development. Consistent with the view that these alterations in canopy patterns are a component of SAS, our results demonstrate they are also influenced by plant density.

Quantitative and Population Genetics: Tools and Application

Impact of genomic correlation between populations on genomic prediction accuracy involving multiple populations

Afees Ajasa[1,2], Solomon Boison[3], Hans Magnus Gjøen[2] and Marie Lillehammer[1]

1 Nofima (Norwegian institute of Food, Fisheries and Aquaculture research), PO Box 210, N-1432 Ås, Norway
2 Department of Animal and Aquacultural Sciences, Norwegian University of Life Sciences, N-1430 Ås, Norway
3 Mowi Genetics AS, Sandviksboder 77AB, Bergen, Norway

The size of the reference population is crucial for the accuracy of genomic prediction. In Atlantic salmon breeding populations, the generation interval is usually three or four years, and there will therefore at any time keep three to four parallel populations (year-classes) in order to produce the needed fish stock for the Atlantic salmon industry. A logical way to increase the reference population size in Atlantic salmon may thus be to combine multiple year-classes. Using gill score records of fish infected with amoebic gill disease (AGD), we investigated the genomic prediction accuracy when combining three year-classes of Atlantic salmon. Three scenarios were investigated, of which two were multi-year-class prediction (evaluated year-class was included in reference set) while the third scenario was an across year-class prediction (excluding the evaluated year-class in the reference set). The difference between the two multi-year-class scenarios studied is that one utilized a genomic based best unbiased prediction (GBLUP) model while the other utilized a multi-trait GBLUP model which allowed genomic correlation (rg, same trait measured in different year-classes) to differ from 1. The consistency of linkage disequilibrium (LD) phase between adjacent markers across year-classes was also studied. The rg between year-classes was generally high (0.99, 0.65 and 0.95 for YC2016 and YC2017, YC2016 and YC2018, and YC2017 and YC2018, respectively), yet the consistency of LD phase between adjacent markers across year-classes was low (0.38 - 0.45), indicating that rg estimates between year-classes are not a good indicator of consistency of LD phase across year-classes. The within year-class accuracy was 0.56, 0.73, and 0.76 for year-class 2016, 2017, and 2018, respectively. The multi-year-class prediction accuracy scenarios studied resulted in no increase in prediction accuracy when compared to the within year-class prediction. The across year-class scenario resulted in relatively low (0.29 - 0.44) accuracy, somewhat proportional to the rg estimates between year-classes, and also very biased. Our results indicate that in a multi-year-class prediction setting, a high rg between year-classes does not mean that these year-classes should be combined as one reference set because the LD phase between markers and quantitative trait loci may differ.

Partitioning of the genetic trend of dairy sheep in Mendelian samplings

Simona Antonios[1], Andres Legarra[1], Riccardo Pong-Wong[2], Jean Michel Astruc[3], Silvia T. Rodríguez-Ramilo[1] and Zulma G. Vitezica[1]

1 GenPhySE, INPT, INRAE, ENVT, F-31326, Castanet Tolosan, France
2 The Roslin Institute and R(D)SVS, University of Edinburgh, Easter Bush Campus,Edinburgh, EH25 9RG, UK
3 Institut de l'Elevage, 149 rue de Bercy, F-75595 Paris, France

This study decomposed the genetic gain in Mendelian samplings by categories of animals defined by sex and by selection pathways. The genetic trend of milk yield for four French dairy sheep breeds (Lacaune, Basco-Bearnaise, Manech Tête Noire, and Manech Tête Rousse) was partitioned. Five categories were defined as: 1) AI males (after the progeny testing), 2) males discarded after progeny testing, 3) natural mating males, 4) dams of males and 5) dams of females. Dams of males and AI males were the most important sources of genetic progress as observed in the decomposition in Mendelian sampling trends. The yearly contributions were more erratic for AI males than for dams of males as they are averaged across a smaller number of individuals. Natural mating males and discarded males did not contribute to the trend in terms of Mendelian sampling as their estimated Mendelian sampling term is either null (natural mating males) or negative (discarded males). Overall, in terms of Mendelian sampling, females contributed more than males to the total genetic gain, and we interpret that this is because females constitute a larger pool of genetic diversity. In this study, the Mendelian sampling term was observed to be the most important factor to determine the selection of a female to become dam, and to maintain the genetic contributions over time.

Characterization and localization of, within, and across breeds as indications of inbreeding depression and heterosCharacterization and localization of, within, and across breeds as indications of inbreeding depression and heterosis

Hafedh Ben Zaabza[1,2], Mahesh Neupane[2], Mohd Jaafar[3], Krishnamoorthy Srikanth[3], Stephanie McKay[1], Asha Miles[2], Heather J. Huson[3], Ismo Strandén[4], Harvey Blackburn[5], Curtis P. Van Tassell[2]

1 Department of Animal and Veterinary Sciences, University of Vermont, Burlington, VT, 05405
2 Animal Genomics and Improvement, Agricultural Research Service, US Department of Agriculture, Beltsville, MD 20705
3 Department of Animal Science, Cornell University, Ithaca, NY 14853
4 Natural Resources Institute Finland (Luke), FI-31600 Jokioinen, Finland
5 National Center for Genetic Resources Preservation, Fort Collins, CO 80521

With the global adoption of genomic selection, the level of inbreeding in dairy cattle has rapidly increased. Inbreeding raises homozygosity throughout the genome and can lead to inbreeding depression in fitness traits. Characterization of dominance effects improves our understanding of the genetic architecture of inbreeding depression. Runs of homozygosity (ROH) are considered to provide an effective measure of genomic inbreeding. The goal in this effort is to characterize the genomic landscape of regions associated with additive, dominance, and ROH effects in 2 breeds of US dairy cattle. We carried out a single-SNP genome-wide association study (GWAS) analysis, fitting SNPs as fixed effects for additive, dominance, and ROH effects one locus at a time. We used a small dataset containing 180,437 pedigreed and genotyped cows born between 2015 and 2020. Genotypes were imputed to approximately 78k SNPs. The analysis was conducted only on yield deviation for 305-d milk yield in Holsteins, but other low heritably traits such as fertility and health will also be analyzed. We found that the most pronounced additive effects were located on chromosomes 14 between 0 and 2Mbp and on chromosome 6 between 20 and 30Mbp. Dominance effects had less pronounced -log10(P) peaks compared to additive effects, and only a few significant dominance effects were detected. The most notable dominance effect was observed on chromosome 5. One region on chromosome 14 also had a large dominance effect. Similarly, ROH effects revealed less prominent -log10(P) peaks than additive effects, although higher than dominance effects. In fact, some narrow ROH peaks were found on chromosome 14. Chromosome 21 also had one region with a large ROH effect. The present study will be further extended to a genomic dataset comprising over 4 million pedigreed and genotyped purebred Holstein cows, nearly 1 million pedigreed and genotyped Jersey cows, and all available crossbreds.

Large-scale single-step genomewide association studies with the algorithm for proven and young

Matias Bermann[1], Daniela Lourenco[1], Ignacy Misztal[1]

1 Department of Animal and Dairy Science, University of Georgia, 425 River Road, 30602, Athens, GA, USA

Genomic selection boosted the genetic progress in many plant and animal populations. The most common statistical model for GS in plant breeding is the so-called genomic best linear unbiased predictor (GBLUP). In livestock populations, where not all the animals are genotyped, the most popular statistical model is single-step GBLUP (ssGBLUP). When the population is large, with tens of millions of animals with records and millions of genotyped animals, performing GS with ssGBLUP is computationally unfeasible. The Algorithm for Proven and Young (APY) was developed to solve this problem by partitioning the individuals into a core and non-core group. The equivalence between (ss)GBLUP and marker-effect models allows doing GWAS using the solutions from routine genetic evaluations. However, such an equivalence vanishes when using APY, and doing GWAS with solutions calculated with APY is theoretically unsound. Using large data sets for GS and GWAS is attractive to increase the statistical power and decrease the false discovery rate. Also, it is desirable to use the solutions from genetic evaluations to perform GWAS for practical reasons. Thus, the objectives of this study were to extend APY to marker-effects models and to derive equations to calculate marker effects and their variances when APY is used to estimate the breeding values.

We derived a family of marker-effects models called APY-(ss)SNP-BLUP. In these models, the genotypes of the non-core individuals are expressed as a linear combination of the genotypes of the core individuals plus an error term. The variance of the marker effect estimates depends on the prediction error variance (PEV) for core individuals but not directly on the PEV of the non-core individuals. With these formulas, performing GWAS using solutions calculated with APY is statistically meaningful. Since all the computations are of dimension equal to the number of core individuals (around 5000-25000 individuals), the proposed formulas are computationally efficient. This computational efficiency allows running GWAS and calculating the accuracy of direct genomic values with large data sets in less than one day, giving a practical solution to increase the statistical power and decrease the false discovery rate of GWAS.

Genetic mapping in interconnected hexaploid sweet potato populations

Gabriel S. Gesteira[1], Marcelo Mollinari[1], Guilherme Da Silva Pereira[2], Bode Adebowale Olukolu[3], Bonny Michael Oloka[1], G. Craig Yencho[1] and Zhao-Bang Zeng[1]

1 North Carolina State University, Raleigh, NC, USA.
2 Federal University of Viçosa, Viçosa, MG, Brazil.
3 University of Tennessee, Knoxville, TN, USA.

Polyploidy is widespread worldwide and present in several economically important crops, such as blueberries, potatoes, sweetpotatoes, grasses, ornamentals, and sugarcane. Due to their high genetic complexity, considering the multiallelic nature in genetic models for polyploids is highly desirable, especially for species that present high ploidy levels and complex population structures. To assess the haplotype composition of such individuals, we developed a genetic mapping procedure that uses an extended hidden Markov chain to recover the multiallelic information on outcrossing populations with multiple founders and any odd ploidy level. We coupled it with a random-effect linear mixed model that takes the haplotype probabilities to perform multiple quantitative trait loci (QTL) mapping in complex traits. We applied our approach to a hexaploid sweetpotato population composed of 703 individuals originated from crosses between three founders. We constructed a genetic linkage map that comprised 28968 variants distributed across the 15 sweetpotato homology groups, spanning 2829 cM of the sweetpotato genome. With the posterior haplotype probabilities, we performed multiple QTL mapping and were able to detect significant QTL for beta-carotene, dry matter, and starch content, with consistent allele effects across sub-populations. The proposed framework provides a good strategy to trace back the founders' alleles to individuals, and increase the genetic informativeness and the statistical power associated with QTL detection. Several outcrossing diploid and polyploid species can also benefit from the proposed framework.

Increasing the predictive ability of maize grain yield under complex scenarios by modeling the additive, dominance, and genotype-by-environment interaction effect and by using the environmental covariates information

Jenifer Camila Godoy dos Santos[1], Natalia De Leon Gatti[2], Antonio Augusto Franco Garcia[3]

1 Nofima (Norwegian institute of Food, Fisheries and Aquaculture research), PO Box 210, N-1432 Ås, Norway
2 Department of Animal and Aquacultural Sciences, Norwegian University of Life Sciences, N-1430 Ås, Norway
3 Mowi Genetics AS, Sandviksboder 77AB, Bergen, Norway

Predictive accuracy of genomic selection (GS) models is a crucial measure because its magnitude will provide the ability of method to predict phenotypic using only genotypic data. Marker density, training population size, genetic architecture of traits, and distribution of linkage disequilibrium are the main factors influencing the predictive accuracy of GS models. However, there are other elements which are little explored but that are also important for the increase of accuracy of models. For example, the dominance effect is essential for accurate genotype prediction, mainly in species like maize which present high levels of heterosis. In addition, the inclusion of Genotype-by-Environment Interaction (GEI) and environmental covariates are also important information that can be used in GS studies. In this context, the purpose of this research is to develop GS models that simultaneously incorporate additive, dominance and GEI effects for predicting maize hybrid's performance in non-evaluated environments, using information from SNP markers and environmental covariates. For that, we are using grain yield data from 156 maize hybrids evaluated in two different crop seasons, at nine locations and under two management conditions in Brazil. Of the 156 hybrids, 149 are single-crosses, represented by three heterotic groups, dent (64 lines), flint (77 lines), and the so-called group C (3 lines). The other hybrids include four common checks (commercial maize cultivars), one double-cross, and two three-way crosses. DNA from the 144 maize inbred lines, used as progenitors of the single-cross hybrids, was extracted from young leaves and genotyped using the standard protocol of Genotyping-by-Sequencing (GBS). We inferred the hybrids' genotypes based on their parents' genotypes using single-nucleotide polymorphism markers obtained via GBS. Finally, we collected the information of ten environmental covariates using the envRtype package. The research is still in progress, and we will have the first results soon. As far as we know, there are still no GS models that use all this information at the same time. Therefore, developing GS models incorporating all effects will significantly contribute to breeding programs, both by theoretical development and practical results.

Development and utilisation of genomic resources and technologies for livestock (dairy) development in Rwanda

***Oluyinka Opoola[1], Felicien Shumbusho[2], Raphael Mrode[3,4], David Hambrook[5], Mizeck G.G. Chagunda[6], Dominic Moran[7], Appolinaire Djikeng[1]**

1 Centre for Tropical Livestock Genetics and Health, The Roslin Institute, University of Edinburgh, Easter Bush Campus, Edinburgh, UK;
2 Rwanda Agriculture and Animal Resources Development Board (RAB), Kigali, Rwanda
3 Scotlands' Rural College, Roslin Institute Building, Edinburgh, UK
4 International Livestock Research Institute, Nairobi, Kenya
5 Royal Jersey Agricultural & Horticultural Society, Trinity, Jersey;
6 Animal Breeding and Husbandry in the Tropics and Subtropics, University of Hohenheim, Stuttgart, Germany
7 Global Academy of Agriculture and Food Security, University of Edinburgh, Edinburgh, UK

*Corresponding and presenting author: oluyinka.opoola@ctlgh.org

In low and middle income countries (LMICs), livestock development is a critical contributor for delivering on the UN sustainable development goals. The productivity and sustainability levels of livestock systems in LMICs remain significantly low compared to the systems in advanced economies. Genomic technologies combined with effective breeding systems can contribute to reducing these gaps and more importantly support the development of context-specific tools and solutions for use to transform livestock production systems in LMICs. The anticipated transformed production systems in LMICs would be characterised by optimum productivity, resilient and healthy animals with an overall reduced impact on the environment.

Our efforts are focusing on the development and harnessing genomic resources and technologies to support Rwanda's long-term investment and strategic orientation to transform the nation's dairy sector. To create the foundational genomic resources, a representative dairy cattle population (n = 2,229) in Rwanda was sampled (by collecting biological samples and associated individual animal metadata including performance recording) and genotyped using the Geneseek Genomic Profiler (GGP, Geneseek® Neogen) 50K SNP array (n = 1,917) and GGP 100K SNP array (n = 312). A combined dataset was subject to quality control, data curation and used in population genetics and genomics analyses. We first assessed the country's dairy cattle population structure, diversity and ancestry. Global dairy populations of European taurine, African indicus and Zebu breeds (n = 250) were used as reference populations to establish the following observations and conclusions.

(continue to page 76)

(continued from page 75)

A subset of our results confirmed that the Rwanda cattle population is a highly admixed population of pure and crossbred animals. This inherent diversity offers the opportunity in a short and long term to breed for adaptation traits, disease tolerance, heat tolerance, productive, profitable and climate-smart animals to inform future selection of desirable breed traits in dairy development.

At the AGBT conference, additional results and conclusions from our initiative would be discussed in the context of sustainable livestock systems for Rwanda and in the much-needed strategies imperatively required to transform livestock-driven food systems while contributing to the global climate change challenges.

VIEWpoly: an interactive tool for mining results from genetic mapping analysis in autopolyploid species

Cristiane Hayumi Taniguti[1], Gabriel Siqueira de Gesteira[2], Jeekin Lau[1], Guilherme da Silva Pereira[3], David Byrne[1], Zhao-Bang Zeng[2], Marcelo Mollinari[2], and Oscar Riera-Lizarazu[1]

1 Texas A&M University, Department of Horticultural Sciences, College Station, Texas, United States
2 North Carolina State University, Bioinformatics Research Center, Department of Horticultural Sciences, Raleigh, North Carolina, United States
3 Federal University of Viçosa, Department of Agronomy, Viçosa, Minas Gerais, Brazil

The genomic complexity of polyploid species required the development of specific and robust methods for genetic analysis. New computational tools allow the construction of linkage maps, infer haplotype transmission, and QTL discovery for many important polyploid crops including potatoes, sweet potatoes, blueberries, forages, roses, and chrysanthemums. Despite the efficiency of the methods, the complexity of the analysis can bring challenges in interpreting results efficiently, connecting them to other sources of genomic information, and relating them with potential plant breeding applications. These challenges can hold back researchers and breeders from making the best use of these advanced new tools. VIEWpoly is a R package for visualizing and exploring results from polyploid computational tools using a shiny interactive graphical user interface. The first version allowed users to upload output files from polymapR, MAPpoly, diaQTL, polyqtlR and QTLpoly and genomic assembly and annotation files. The results are displayed in three different modules: i) VIEWqtl: shows the results of QTL analysis, including significance profile, confidence intervals, estimated effects, a table with breeding values, and an interactive search for individuals containing specific haplotypes; ii) VIEWgenome: shows the relationship between the linkage map and QTL analysis with the reference genome sequence and annotated genome making use of the JBrowseR interface. iii) VIEWmap: displays the SNP dosages in parentals and offspring, and their respective haplotypes in the genetic map. All modules provide features to download specific information into comprehensive tables for further analysis and presentation. The modular structure of VIEWpoly facilitates the expansion of the software to include results from other software and analysis. In its second version, VIEWpoly provides support for diploid and multi-population analysis results. We also released directions and a VIEWpoly GitHub branch to help users make their results publicly available using the VIEWpoly interface. VIEWpoly's user-friendly interface highlights the capabilities of the implemented tools and facilitates the efficient use of the results even by people without R programming skills.

Index

Index (continued)

Index (continued)