



## 2020 G2F GxE Collaborator's Meeting Phenome Meeting

**Date:** Monday, February 24<sup>th</sup>, 2020

**Time:** 11:15AM to 4:00 PM

**Location:** Tucson Convention Center, Maricopa Room I, 260 S Church Ave, Tucson, AZ 85701

### Agenda

**11:15 – 11:25AM – Welcome and Introductions**

#### **G2F Updates:**

11:25 – 11:40AM – G2F Description & 2019 Data Summary – Alejandro Castro Aviles, Univ. of Wisconsin

11:40 – 11:55AM – G2F Funding Status & Lobbying Efforts – David Ertl, Iowa Corn Growers

11:55 – 12:10PM – 2020/21 Plan Summary & Future Activities – Natalia de Leon, Univ. of Wisconsin

**12:10 – 12:40PM – *Break and executive lunch served***

#### **Research Presentations:**

12:40 – 1:05PM – “Plugin”-Based Architecture of Software to Predict Corn Phenotypes – Francisco Munoz-Arriola, University of Nebraska, Lincoln

1:05 – 1:30PM – Prediction of Complex Maize Traits Using Machine Learning Methods that Incorporate Genomic and Environmental Covariates - Cathy Jubin, Georg-August-Universität Göttingen

1:30 – 1:55PM – Characterizing Genetic and Environmental Impacts on Maize Using Phenomic Approaches – Sara Tirado, Univ of Minnesota

1:55 – 2:20PM – Computational Approaches to Analyze Big Root Data Grown in the Field - Alexander Bucksch, University of Georgia

2:20 – 2:45PM – *coffee break*



2:45 – 3:10PM – Validation of Functional Polymorphisms Affecting Maize Plant Height by Unoccupied Aerial Systems (UASs) allows Novel Temporal Detection - Alper Adak, Texas A&M University

3:10 – 3:35PM – Semantic Segmentation of Sorghum Using Hyperspectral Data Identifies Genetic Associations for Morphological Traits - Chenyong Miao, University of Nebraska, Lincoln

3:35 – 4:00PM – Managing UAS Imagery for Developmentally-Driven Decision Making and Genetic Analysis - Katy Rainey, Purdue University

4:00PM - Adjourn



Title: “Plugin”-Based Architecture of Software to Predict Corn Phenotypes

Munoz-Arriola, F., D. Jarquin, H. Hohbein, P. Sarzaeim, J. Carter, D. Recic, Z. Trautman, A. Zhang, C. Espinoza-Garcia, G. Lopez-Morteo, B Ramamurthy

Departments of Biological Systems Engineering and the School of Natural Resources, Institute of Agriculture and Natural Resources, University of Nebraska-Lincoln

ABSTRACT:

Scientific and engineering actions are taking place to better understand agroecosystems responses to environmental changes. Some of the responses to those changes are evidenced in the genetic composition and phenotypes of a single plant or a community of plants resisting droughts or pests. Here, we consider that applied research driven by farmers’ knowledge and data available and basic research articulated in theoretical foundations of genetics and the environment play a major role in sustaining the intensification of agriculture. The **goal** of this work is *to present an architecture of software aimed to collect and integrate multidimensional data from multiple sources to predict crops’ responses to hydroclimatic changes*. The platform is built upon an AWS Elastic Beanstalk in order to deploy and scale the Dataplugin –a software coded in Java that integrates multidimensional (genetic, environmental, and climate) data. The data collection is selective and can be retrieved from multiple sources but treated as single clients. Predictive data analytics for machine learning and GxE -coded in python and R— deployed and scaled. Preliminary results include backend and frontend, off-line machine learning approaches for database improvement and GxE model for prediction of phenotypes.



Title: Prediction of Complex Maize Traits Using Machine Learning Methods that Incorporate Genomic and Environmental Covariates.

Authors: Cathy Jubin<sup>1</sup>, Henner Simianer<sup>1,2</sup>, Timothy Mathes Beissinger<sup>1,3</sup>

1 Center for Integrated Breeding Research, University of Goettingen, Germany

2 Animal Breeding and Genetics Group, Department of Animal Sciences, University of Goettingen, Germany

3 Division of Plant Breeding Methodology, Department of Crop Sciences, University of Goettingen, Germany

#### ABSTRACT:

Crop performance depends on complex interactions between genetic background and environmental conditions. Weather-related covariates can be used to characterize environments in plant breeding trials. Machine learning techniques may be capable of identifying the most relevant environmental variables that impact important traits. Additionally, machine learning approaches sometimes enhance the ability to predict phenotypes, compared to linear models, in multi-environment datasets. To explore the promise of machine learning for identifying critical weather-related variables and using this information to predict complex traits in maize hybrids, we leveraged the publicly available resources of the Maize Genomes to Fields (G2F) Initiative from 2014 to 2018. We developed an assortment of basic environmental variables covering four maize growth stages: early vegetative, late vegetative, flowering, and grain filling. Additional environmental variables that exploited prior knowledge about crop physiology were also used. As expected, many variables were highly correlated, such as heat stress covariates with minimal, maximal and average temperatures. Therefore, feature (referred to as variable) selection methods were compared in order to obtain a subset of the most relevant variables to use as covariates in our modeling. Recursive feature elimination and Lasso regularization identified common environmental variables explaining most of the trait variability, for instance heat stress at flowering stage. In addition, GBS data of inbred lines were processed to obtain synthetic genotypic data (250,000 SNPs remaining for further analyses) on approximately 2,200 maize hybrids. Our initial results demonstrate that models with environmental covariates can sometimes outperform models which utilize year-location covariates for prediction. In this talk, we further dissect predictability by describing the potential of machine learning methodologies for feature selection and hybrid prediction.



Title: UAV Based Imaging Platform for Characterizing Genetic and Environmental Impacts on Maize Growth

Sara B. Tirado<sup>1,2</sup>, Candice N. Hirsch<sup>1</sup> and Nathan M. Springer<sup>2</sup>

1 Department of Agronomy and Plant Genetics, University of Minnesota, Saint Paul, MN 55108

2 Department of Plant and Microbial Biology, University of Minnesota, Saint Paul, MN 55108

#### ABSTRACT

Plant height (PH) data collected at high temporal resolutions can give insight into how genotype and environmental variation influence plant growth. However, in order to increase the temporal resolution of PH data collection, more robust, rapid and low-cost methods are needed to evaluate field plots than those currently available. Due to their low cost and high functionality, unmanned aerial vehicles (UAVs) provide an efficient means for collecting height at various stages throughout development. We have developed a procedure for utilizing structure from motion algorithms to collect PH from RGB drone imagery and have used this platform to characterize a yield trial consisting of 24 maize hybrids planted in replicate under two dates and three planting densities planted in St Paul in the summers of 2018 and 2019. Plant height data was collected using both weekly UAV flights and manual measurements. The comparisons of UAV-based and manually acquired PH measurements revealed sources of error in measuring plant height and were used to develop a robust pipeline for generating UAV-based PH estimates. This pipeline was utilized to document differences in the rate of growth between genotypes and planting dates. Our results also demonstrate that growth rates generated by PH measurements collected at multiple timepoints early in development can be useful in improving predictions of PH at the end of the season. We also document the ability to use UAV PH estimates to assess variation for lodging responses for plants at different planting densities or developmental stages. This method provides a low cost, high throughput method for evaluating plant growth in response to environmental stimuli on a plot basis that can be implemented at the scale of a breeding program.



Title: Computational Approaches to Analyze Big Root Data Grown in the Field

Alexander Bucksch

Department of Plant Biology, Institute of Bioinformatics, Warnell School of Forestry and Natural Resources, University of Georgia

ABSTRACT:

The use of mathematical and computational approaches within the plant sciences enables two new major areas: The discovery of new phenomena in existing data and detailing phenotypic measurements beyond the manually achievable. In particular, phenotyping generates large amounts of imaging data at very high resolutions. Both, increased data volume and physical resolution of root architecture demanded the development of parameter-free algorithms that will process 3D imaging data without user input on high-performance computing systems. The coupling of high-performance computing with image acquisition systems will allow for the discovery of uncharacterized phenotypes that were previously “invisible” because of low sampling rates. The talk at the Genome2Fields workshop will therefore highlight technological advances developed at UGA to collect 3D root architecture data with unprecedented detail for field grown roots. The presented advances include a 3D root scanner using a newly developed structure-from-motion algorithm to resolve occlusions within the root system. The talk will also demonstrate the validation of new 3D algorithms to compute root system and individual root traits as well as a vision to analyze existing big data of roots to reveal uncharacterized phenomena.

Title: Validation of Functional Polymorphisms Affecting Maize Plant Height by Unoccupied Aerial Systems (UASs) allows Novel Temporal Detection

Alper Adak, Seth C. Murray, Clarissa Conrad, Yuanyuan Chen, Nithya Subramanian, Steven Anderson, Scott Wilde

ABSTRACT:

The use of UAS phenotyping can provide functional information about the phenotypic effect of single nucleotide polymorphisms (SNP) on complex trait (e.g. plant height) using advanced breeding populations. Two SNPs affecting plant height (PHT) and grain yield in maize (*Zea mays* L.) were previously discovered using diversity panel of maize hybrids in GWAS study in Texas and have been implemented into Kompetitive Allele Specific PCR (KASP) assays. Backcross-based heterogeneous inbred families (HIFs) were advanced using four different linkage mapping populations during 2016 to 2019 using the two SNPs to (i) advance the HIF populations and (ii) validate allelic effect sizes of the SNPs on PHT during multiple growth stages via UAS surveys. Weekly PHT measurements were estimated from UAS survey images from emergence to harvest (5/17/19 to 7/12/19). Manual PHT measurements were collected at a single time point nearing harvest (7/02/2019), as typically done. Manual PHT measurements results showed that the best linear unbiased estimations (BLUEs) of the dominant allele (XX calls) contributed by NC356 and LH82 were consistently 2 to 12 cm taller, depending on the genetic background of each HIF. UAS PHT estimates confirmed hand measurements ( $r=0.85$ ), and for the first time, demonstrated a temporal dimension to these SNP's allelic effects size. In early growing stages the allelic effect sizes of each SNPs varied from 10 to 14 cm narrowing to 4 cm as populations reached termination of terminal growth stages, when manual PHT measurements are typically taken. Our results demonstrated that allelic effect sizes of quantitative traits can be dynamic in temporal growth resulting in informative phenotypic variability is being overlooked following traditional phenotyping methods.



Title: Semantic Segmentation of Sorghum Using Hyperspectral Data Identifies Genetic Associations for Morphological Traits

Chenyong Miao and James Schnable

Department of Agronomy and Horticulture, Center for Plant Science Innovation, University of Nebraska, Lincoln

ABSTRACT:

This study describes the evaluation of a range of approaches to semantic segmentation of hyperspectral images of sorghum plants, classifying each pixel as either nonplant or belonging to one of the three organ types (leaf, stalk, panicle). While many current methods for segmentation focus on separating plant pixels from background, organ-specific segmentation makes it feasible to measure a wider range of plant properties. Manually scored training data for a set of hyperspectral images collected from a sorghum association population was used to train and evaluate a set of supervised classification models. Many algorithms show acceptable accuracy for this classification task. Algorithms trained on sorghum data are able to accurately classify maize leaves and stalks, but fail to accurately classify maize reproductive organs which are not directly equivalent to sorghum panicles. Trait measurements extracted from semantic segmentation of sorghum organs can be used to identify both genes known to be controlling variation in a previously measured phenotypes (e.g., panicle size and plant height) as well as identify signals for genes controlling traits not previously quantified in this population (e.g., stalk/leaf ratio). Organ level semantic segmentation provides opportunities to identify genes controlling variation in a wide range of morphological phenotypes in sorghum, maize, and other related grain crops.