2018 NIFA FACT Workshop

High Throughput, Field-Based Phenotyping Technologies for the Genomes to Fields (G2F) Initiative

January 28-30, 2018

Ames, IA USA

Organizers
Patrick Schnable, *Iowa State University*
Natalia de Leon, *University of Wisconsin, Madison*
David Ertl, *Iowa Corn Growers Association*
Nathan Springer, *The University of Minnesota*

Authors
Carolyn J. Lawrence-Dill, Patrick Schnable, Nathan Springer, and:
Natalia De Leon, Jode Edwards, David Ertl, Shawn Kaeppler, Nick Lauter, John McKay, Francisco Munoz-Arriola, Seth Murray, Duke Pauli, Nathalia Penna Cruzato, Colby Ratcliff, James Schnable, Kevin Silverstein, Edgar P. Spalding, Addie Thompson, Ruth Wagner, Jason Wallace, Justin Walley, and Jianming Yu.

Acknowledgements
These students, postdocs, staff members, and industry scientists participated in discussions formative for the guidance and suggestions reported: AJ Ackerman, Naser Alkhalifah, Steven Anderson, Matt Baseggio, Jessica Bubert, Darwin Campbell, Celeste Falcon, Matt Johnson, Aaron Kusmec, Bridget McFarland, James McNellie, Danilo Moreta, Stefan Hey, Amritpal Singh, and Adam Vanous.

This workshop was sponsored by the United States Department of Agriculture – National Institute of Food and Agriculture, Food and Agriculture Cyberinformatics and Tools (FACT) program via grant number 2017-07799 P.S.S. with additional support provided by Iowa State University's Plant Sciences Institute. Any opinions, findings, conclusions, or recommendations expressed in this report are those of the participants and do not necessarily represent the official views, opinions, or policy of the funding agency.

Executive Summary

The U.S. has long played a leading role in developing advanced agricultural technology. We must continue to develop novel approaches to increase the production of food, fiber and fuel while protecting our natural resources and ensuring an economically vibrant agriculture sector. The Genomes to Fields (G2F) Initiative strives to take advantage of new technologies to improve the productivity and stability of maize. In particular, this initiative seeks to connect advances in our understanding of crop genomes with new robotics, high-throughput sensing technologies, and other data gathering devices to understand how plant traits are influenced by genetics and the environment with a long-term aim of developing crops that will exhibit sustainable enhanced productivity across diverse sets of environments and years.

In January 2018, G2F hosted a meeting in Ames, Iowa to consider trajectories for research in field-based phenotyping and how best to support those trajectories, and informing USDA NIFA and other federal agencies of our findings and conclusions.

For four years, a distributed network of G2F field sites and public-sector collaborators has been generating the data needed to develop predictive models. Generating a vibrant community of researchers from diverse disciplines including the breadth of the plant sciences (e.g., genetics, agronomy, physiology, modeling, and breeding), engineering, computational sciences, and climatology is critical to the full success of the initiative.

Many topics were discussed over the course of the three-day workshop, but four themes emerged as areas ripe for focused effort in the coming years: (1) support for ongoing **community experiments**; (2) development and use of **field sensors and plant imaging platforms**, (3) creation of **data management, sharing, and analytics platforms** especially for making effective use of large scale image sets, and (4) **engaging additional scientific disciplines in G2F and training the next generation of agricultural scientists**. It was suggested that these topics were of interest to advance not only the goals of the G2F initiative, but also the field of predictive plant phenomics more generally. The group supported the idea to enhance coordination efforts across all four themes by designating and/or creating a number of (5) **High-Intensity Phenotyping Sites (HIPS)** where individualized research areas and local expertise could develop alongside intentional, coordinated linkages focused on advancing topics within the four shared themes. While community experiments are necessarily extensive, these HIPS experiments can be embedded within the G2F testing network and allow for more intensive investigation and development of predictive phenomics tools. Tools and learnings from the HIPS could then be deployed more broadly.

Background: Ensuring a secure supply of agricultural products for nutritious food, feed, fiber, and fuel is a significant short- and long-term challenge for US agriculture. G2F is an umbrella initiative to support the translation of maize genomic information for the benefit of farmers, consumers, and society. It builds on publicly funded maize genome sequencing projects to develop approaches to understand the functions of maize genes and specific alleles across environments. Ultimately, this information will be used to enable accurate prediction of maize phenotypes in diverse environments and to develop varieties with stable production in farmers' fields. There are many dimensions to the broad goal of understanding genotype-by-environment (GxE) interactions, including identifying genes which impact traits and trait components, how genes interact with other genes (GxG), the relevance of specific alleles under different growing conditions, and how these alleles influence plant growth during various stages of development. The G2F initiative promotes projects that advance integrated research and technologies, combining disciplines such as genetics, genomics, plant breeding, plant physiology, agronomy, climatology and crop modeling, with computation and informatics, statistics and engineering. Since 2014, this project has evaluated more than 66,000 field plots involving more than 1,500 maize hybrid varieties across 115 unique environments in North America. These datasets can be used to construct novel algorithms to analyze breeding data and to develop new breeding methods that increase crop productivity and stability and/or reduce environmental impacts of crop production.

Workshop Activities: A subset of the interdisciplinary group of researchers engaged in the G2F initiative met over the course of three days to discuss novel sensors and other technologies, tools, and approaches for gathering, processing, and sharing G2F data specifically, and other large-scale datasets more generally. The goal of the workshop was to identify field-ready sensors and data analysis mechanisms to start using now as well as to identify areas where investments in sensor development, field equipment, and improved data management would advance the goals of G2F. The workshop involved a reception and keynote address the first evening; seminars, a panel discussion, and a breakout group discussion period on day two; and lightning talks and further breakout group discussion on day three followed by a tour of the Pioneer Hi-Bred high-throughput phenotyping facilities. Aiming to catalyze discussion, speakers delivered live-streamed presentations on broad topics including automated data collection, field management, data analysis, and implications. The panel discussion focused on technology transfer and guidance on how best to initiate start-up companies including both specific guidance on how to begin such endeavors and qualities that personnel involved in startups must possess and develop. Breakout group discussions were broad, but became focused on identifying potential community experiments; field sensors; data management, sharing, and analytics; and broadening participation to diverse disciplines and training the next generation of scientists and engineers. These activities were documented through shared live note-taking, with raw notes available as Workshop Meeting Notes at the G2F Workshop Website (http://www.genomes2fields.org/meetings/2018-nifa-fact/). We highlight here some of the recommendations and discussions that occurred during the workshop breakout groups and community discussions. Based on guidance from NIFA, returning information to the agency in a timely fashion was a priority. For that reason, these recommendations are presented as high-level bullet points and have not been extensively polished.

Findings: Many topics were introduced and discussed over the course of the three-day workshop, but four themes emerged: (1) support for ongoing community experiments, (2) deployment, evaluation, and use of field sensors and imaging technologies, (3) creation of data management, sharing, and analytics platforms, and (4) training the next generation of agricultural researchers.

In the end, it was agreed that all four areas could be advanced not only through continued efforts at current and future G2F project sites, but also through (5) designating and/or creating High-Intensity Phenotyping Sites (HIPS) with individualized areas of expertise but shared goals to focus on these four topics.

(1) Community Experiments:

- Geographically distributed community experiments have great potential to advance our understanding of GxE interactions as demonstrated by the work already conducted (Gage et al., 2017). Inclusion of controlled/managed-stress environments would add substantial value.
- Community experiments must be focused on relevant scientific questions, including not only relevant hypothesis testing-based research but also data-driven hypothesis generation efforts including prediction of phenotypes. Validation of findings and in-depth characterization are key. We seek to understand gene function as well as to impact relevant phenotypes and traits across environments and germplasm.
- Experiment-wide phenotyping will facilitate new interpretations. Piloting new technologies by extensive and systematic phenotyping of a small set of inbreds and hybrids at targeted sites is a useful way to evaluate relative value of technologies and logistical issues that will be faced upon broader deployment.
- Collection of new (more resource-intensive) types of data is envisioned across sites. This is expected to benefit further advancements in data analysis and interpretation. New data types will support new types of analysis and interpretation. Examples include canopy temperature, physiological and developmental attributes (i.e., those needed to parametrize process-based crop growth models), and detailed morphologies (such as stomatal density). Difficult-to-measure phenotypes may be targeted to HIPS, with interpretations or surrogate measures subsequently extended more broadly.
- Significant seed production is needed to support a project of this magnitude. The seed needs of G2F hybrids across locations can exceed 5,000 kernels per year and we should target production of 50,000 kernels of focus hybrids and inbreds for high-intensity analysis. Parent inbred seed also needs to be available for acquisition by external researchers. To facilitate experimental implementation and to reduce year-to-year variability due to seed lot differences, large amounts of specific seeds lots of core G2F germplasm is critical, along with a mechanism to identify and request stocks.
- Different sites could have different goals based on available resources and strengths. High-throughput phenotyping efforts need to be located in sites where accessibility to labor and other infrastructure are conducive to such efforts. Other methodologies and techniques like shovelomics also are a goal.
- Different types of experiments within the larger network would help with engagement of diverse research communities. Experiments that require specialized field equipment will only be possible in sites that have access to such equipment, but other efforts, such as those that might involve engineers and computer science experts, can be accomplished at other sites.

- Experiments within the G2F portfolio should range from quantitative genetic/breeding assessment of GxE including genomic prediction, to physiology and crop modeling, to genetic evaluation of major gene effects. G2F welcomes engagement of a broad research community. Some experiments require specialized planting and harvesting equipment, and others may be able to be conducted by groups without equipment but rich in expertise. The portfolio of cooperators will therefore extend through the entire spectrum from traditional field-based researchers at land-grant institutions to scientists with undergraduate research teams at primarily undergraduate institutions. Computational and engineering activities, for example, could occur at institutions without any capacity for field-based research.

(2) Field Sensors and Plant Imaging Platforms:
We recommend defining and widely deploying a set of shared sensors and other field-ready technologies across all G2F sites. This would be further enabled by standardizing measurements and/or devices.

- Soil monitoring should be linked to weather stations and should involve efforts to:
    - Measure water availability across three different depths. This would involve multiple probes (and/or wells) per location with exact numbers based on location data (some locations would need more dense monitoring than others).
    - Monitor nitrogen similar to plans outlined for water availability.
- Individual plants should be analyzed to enable:
    - Plant transpiration monitoring: Initially deploy multiple sensors per plant on three plants per plot for a subset of inbreds or hybrids (~25 genotypes) at ~12 locations. The information from this set could be used to design further experiments to decide how many sensors per plant, how many plants per plot, when to apply, frequency of collection and how to collect data.
    - Nitrogen monitoring: Combine stable monitoring and mobile monitoring. Do initial phase on subset of inbreds or hybrids to determine the proper location to place sensor, number of sensors per plot and value of time course data.
    - Stationary camera systems to monitor growth and development of plants
- UAVs (Unmanned Aerial Vehicles; also called drones or UAS; Unmanned Aerial Systems)
    - Personnel must monitor changes to regulatory framework and new technologies that could automate flights and data collection.
    - An ideal UAV system would at least generate stand count, plant height, NDVI, canopy closure, flowering, canopy temperature and multi-spectral reflectance on an approximately weekly basis.
- UGVs (Unmanned Ground Vehicles)
    - A primary goal is to capture 3D structures of plants to monitor stem thickness, leaf numbers and angles, plant architecture, flowering date, etc.
    - The initial phase of UGV work within G2F will involve the deployment of the TerraSentia rover (https://www.earthsense.co/home) across three sites in 2018. This unit can carry a variety of sensors and will be initially deployed to measured stand counts, early season plant vigor and indicators of physiological stress.
    - Development phase systems to be deployed include the ARPA-E prototype for 2018 and the NSF- funded PhenoBot3.0 (which is an outgrowth of earlier USDA funding); alternative options.
- Satellite data: Investigate potential to partner with high-resolution satellite providers. Our ground truth data would be an asset in such collaborations.

(3) Data Management, Sharing, and Analytics:

It was specifically noted that a funded, single large project (CAP or BRC type) would enable standardization and coordination that would far exceed any group of small projects. Coordination would lead to a comprehensive, uniform, and annotated data set. We strongly support the large project model for project funding.

Current and future data management, sharing, and analytics for the G2F GxE project were outlined and discussed, with specific changes and suggestions articulated. In particular, it was noted that ingesting data from all sites needs to be streamlined.

Over the course of discussion, it became clear that three areas are critical for success of the effort: improved communication platforms and mechanisms, a community-oriented data access portal, and platforms for data storage.

- We must facilitate online communication among G2F participants. Updates to the existing G2F website are needed to enable people to determine who is working on what projects/data (at least PIs, possibly also postdocs and students). Access lists of G2F-related publications & presentations would also be beneficial.
- A phenomics community data portal would help. Example functionality should include the ability to support multiple data types including soil data, water availability, weather/climate, images (from stationary or mobile platforms), stand count, canopy temperature, NDVI, plant height and other extracted values.
- A large-scale phenotypic data store is needed. It is unclear at this time where to store image-based, remote sensing, and other phenotyping data

For data analytics, the following observations were made:

- Standardized methods are valuable, but novel methods are, too! Without flexibility and support for novel data collection, analysis, and organization, data storage and analysis cannot improve.
- Timely data upload must occur for collaborators to be included in future iterations. The group has engaged with the International AgroInformatics Alliance (IAA) that is part of the Super Computing Institute at the University of Minnesota to enhance automated data curations capabilities and to develop a user-friendly interphase and database infrastructure for data intake starting in 2018. This will facilitate data curation and therefore timeliness of data movement.
- On-site data extraction enables quality checks (requires expertise and commitment to execute).
- Training in equipment operation and data analysis both on-site and via central processing is critical.

(4) Training the Next Generation of Agricultural Scientists:

- Expansion into additional disciplines is encouraged.
    - The set-up of the G2F initiative provides an enormous opportunity to engage members of other disciplines that can make tremendous contributions to the foundational goal of this effort beyond genetics and breeding. Targeted efforts are needed to engage physiologists, biochemists, crop modelers, soil scientists, ecologists, climatologists, engineers, computer scientists, economists, and communication specialists (internal as well as external, including policy-makers).
    - If HIPS (see section 5) focus on an in-depth topic, expansion toward other types of scientists in keeping with the needs of HIPS makes natural sense. Technological and analytical developments generated at the small scale can subsequently be distributed more broadly through the distributed network of phenotyping sites within G2F.

- G2F training needs to occur across many dimensions.
    - G2F has done a good job of communicating with stakeholders (farmers, funding agencies, university administrators, etc.).
    - G2F should develop centralized training opportunities for phenotyping technologies, protocols for data transfer, documentation, etc.
    - G2F should continue to identify and articulate a clearer idea of what specific questions and research objectives we are trying to address. Clear delineation of those goals provides more opportunities to engage members from other disciplines.

- Many students aim to join industry.
    - We need specific programs to be deployed across many universities. It is possible that within the general area of plant phenomics, the programs could have shared curricula. Iowa State University's P3 graduate program that is training students at the intersection of plant science, engineering, and computational sciences might be a useful model.
    - Interdisciplinary research requires time. How long the teaching-learning process will take should be a major consideration.

- Students, postdocs, and professionals need training outside their primary discipline. There is a need for training opportunities and workshops focusing on topics such as:
    - How to participate in team science (i.e., how to effectively collaborate with others outside of their primary discipline).
    - How to deploy sensors and collect, aggregate, and analyze data

  To address these needs, training workshops and symposia are recommended.

(5) High-Intensity Phenotyping Sites (HIPS) are needed to coordinate across the other four dimensions

The G2F initiative continues to strive to narrow the gap to move from proof-of-concept projects to enabling widespread use of new sensors and phenotyping devices at a broader scale. The G2F Initiative will need to make use of new measurement technologies across many locations. Engineering colleagues have developed new technologies to enable measurement of plants, soil and fields and it is important to develop routes for the widespread adoption of these technologies across many sites. It is important to note that true adoption requires utilization of the device/sensors, uploading of data to a common location and processing of data to extract traits of interest. One concept that arose from our workshop was to leverage HIPS as a mechanism to beta-test new technologies before widespread deployment. New technologies would be deployed at HIPS for one or two years to assess and optimize their ease-of-use, accuracy and value prior to widespread adoption at all G2F sites.

A set of 3-6 locations would be designated as HIPS according to a set of guidelines to be established. Each site would need to be willing to develop appropriate infrastructure for deploying technology/sensors, collecting data, and performing ground-truth measurements of plant traits. If necessary, HIPS would utilize paired field plots to collect yield data for the same genotypes; some sites might also utilize variable planting density, nitrogen availability, planting dates, or water availability to provide information on how the technologies can detect changes in plants due to specific environmental perturbations. A large amount of seed stocks for a selected set of genotypes would be generated and used consistently at the HIPS to ensure standardization of plant materials within and across HIPS and years. The data from HIPS would be made freely available in the same way as the other G2F data and reports from each site would document the technical requirements and ease-of-use for each technology. HIPS would also be a good place to grow specific genetic materials to investigate how specific gene functions relates to traits of interest (via a candidate gene approach, for example).