



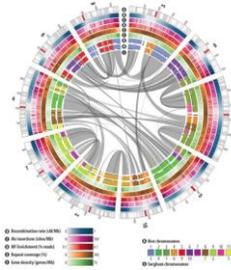
## Genomes to Fields Initiative



Patrick S. Schnable  
Iowa State University  
10 December 2013



## How to Translate Genomic Data into Greater Understanding of Plant Biology and Crop Improvement?



Schnable, Ware et al. (2009) Science

The Maize Genome Sequencing Project, Rick Wilson, PI

Maize contains ~50,000 genes

2

## “Guilt-by-Association” Studies



- The young Teresa Wright character suspects her Uncle Charlie of being the "Merry Widow Murderer".
- Postmarks on letters from him correlated with dates and locations of murders.

- Association studies link SNPs (DNA variation) and genes to traits
- GWAS requires high quality genotyping and phenotyping data

## The Challenge of Phenotyping

“It should be recognized that understanding the plant phenome is as challenging, complex and important, and as deserving of its own methods, approaches and standards, as is understanding the genome”

-Jonathan Lynch, Penn State Univ



\* it will almost certainly be MORE challenging and complex... -PS

## The G2F Concept

- Logical next step to apply information generated by the sequencing of the maize genome and exploration of the genetic diversity of this species for crop improvement
- Coordinated project involving many genotyped lines that are phenotyped at many well characterized locations
  - To assess allele-by-environment interaction with sufficient resolution to enable accurate modeling and prediction of the performance of specific genotypes in specific environments
  - To provide venue for researchers to test creative ideas
- Public-private partnership
- Develop unique phenotyping technologies and analysis methods
- Results to be shared with the entire community
- Important training opportunity

## “Big Data” and the G2F Initiative

- “Big Data encompasses the acquisition, management, analysis and interpretation of information with extreme volume, variety, velocity (rate of acquisition), veracity, variability and complexity” -Carolyn Lawrence
- The Big Data paradigm enables researchers to more quickly formulate and test novel hypotheses.
- G2F Initiation envisions using the approaches of Big Data to develop models that enable the prediction of phenotypes based on genotypes and environmental data



Phenotype (P) = Genotype (G) + Environment (E) + GxE

- Genotype: NGS revolution and GBS
- Environment: weather, soil type, water, nutrients, disease pressure, agronomic practices etc.
- GxE interactions complicate phenotypic predictions, but offer fascinating avenues of investigation



7

E and GxE complicate phenotypic predictions

- Strategies for dealing with “E” and “GxE”
  - Conduct studies in controlled environments, taking E and GxE out of the equation
  - Control for *and study* the effects of E and GxE statistically... *embrace the opportunity to gain a deeper understanding of the underlying biology*

8

## Develop *Predictive* Statistical Models (Netflix Challenge)

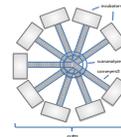
- Establishing a large public-private partnership with seed funding from the ICPB
- Pilot study in progress for 2014
- Phenotype many genotyped lines in many well characterized environments.
- Use appropriate experimental design and statistical analyses to separate the effects of G, E and GxE
- Plan is to scale-up to 1,000s of genotypes at hundreds of locations by summer 2015 and continue for 5-10 years



Natalia DeLeon  
Univ of WI

## Construction and Testing of Models will Iteratively Enhance our Understanding of Maize/Corn Biology

Models will enable us to define hypotheses about the effects of G, E, and GxE on P that can be tested in experiments in which G and E can be tightly controlled (e.g., in controlled environments such as the “Envirotron”). The outcomes of these experiments can be used to iteratively refine predictive models.



## Predictive Models Will:

- Improve the accuracy of selection in plant breeding programs, thereby increasing the rate of genetic gain per year
- Enhance our ability to efficiently breed crops to withstand the increased weather variability associated with global climate change
- Improve the ability of farmers to select appropriate varieties for their growing conditions and management practices

## G2F Initiative Executive Committee

- Pat Schnable (Iowa State Univ), co-lead
- Natalia DeLeon (Univ of WI), co-lead
- Ed Buckler (USDA/Cornell),
- Sherry Flint-Garcia (USDA/Univ of MO)
- Shawn Kaepler (Univ of WI)
- Jonathan Lynch (Penn State Univ)
- Nathan Springer (Univ of MN)
- David Ertl, Iowa Corn Growers' Association
- (*representatives of other disciplines to be added*)

## G2F Initiative Time Line

- Fall 2012 Presented concept to Iowa Corn Promotion Board
- March 2013, G2F Executive committee organized at the Maize Genetics Conference; discussions continued at NCCC167 meeting
- Spring/Summer 2013, engaged selected industry partners
- Fall 2013, organized a workshop to address logistical and technical questions and IP issues
- Dec 2013, community engagement at the ASTA meeting
- Jan 2014, community engagement at the PAG meeting
- Summer 2014, Conduct pilot study (limited environments and genotypes); large-scale seed increase
- Summer 2015, 1<sup>st</sup> large-scale study



## Decadal Plan for the Plant Sciences

- Major report prepared by the ASPB with support from HHMI, NSF, USDA, DOE under the leadership of David Stern and Sally MacKenzie
- First research goal is Predictive Phenomics: "Increase the ability to predict plant traits from genome in diverse environments (lab, field, and nature)"
- The G2P WG of the NSF's iPlant Collaborative shares this goal



## Vision for Federal Agencies, Congress and Industry

- Apply information generated by the sequencing of the maize genome and exploration of the genetic diversity of this species to crop improvement
- Will engage a broad spectrum of the scientific community (both directly and indirectly via access to data)-ecosystem model (human genome project)
- Initiative will be expensive: multiple "projects" within the initiative, likely with multiple, likely discrete funding streams
- Public-private partnership: collaborative effort in drafting science plan
- As was the case with the maize genome sequencing project there is strong support from the Corn Growers
- Goals include both enhanced ability to predict plant performance and a deeper understanding of relevant biology
- Impacts both breeding and basic plant biology
- Important training activity in the "Big Data" era