



<http://www.genomes2fields.org>

# Genomes to Fields

## Data Management and Analysis Solutions for Maize Predictive Phenomics: A partnership with the GxE Subgroup of the Genomes to Fields (G2F) Initiative

**GxE Data Management Team<sup>1</sup>**  
**Genomes to Fields Consortium<sup>1</sup>,**  
**Jack M. Gardiner<sup>2</sup>, Darwin A. Campbell<sup>2</sup>, Ramona Walls<sup>3</sup>, Jeremy de Barry<sup>3</sup>, Matthew Berrigan<sup>4</sup>, and Carolyn J. Lawrence-Dill<sup>2,5</sup>**

<sup>1</sup>See Consortium Participants, <sup>2</sup>Department of Genetics Development and Cell Biology, Iowa State University, Ames IA, <sup>3</sup>iPlant Collaborative, University of Arizona, Tucson AZ, <sup>4</sup>Leafnode, Auckland New Zealand, <sup>5</sup>Department of Agronomy, Iowa State University, Ames IA

### Abstract

Breeding decisions to develop improved cultivars for industrial and agricultural uses are greatly facilitated by simultaneously leveraging phenotypic, genotypic, weather, and image data. Development of standardized data collection and analysis practices by working directly with the data generators in the initial stages of the data collection process, is an important first step to support a multi-institutional, multi-year breeding projects. The **Genotype by Environment (GxE)** subproject within the maize **Genomes to Fields (G2F)** is a multi-institutional project spanning 25 North American locations. The GxE group is collectively addressing environmental effects on the performance of a large collection of maize inbreds and hybrids grown in 25 diverse locations. Each location is collecting data on 14 core phenotypic traits, as well as weather measurements with image data for a subset of locations. To assist in the management of these diverse data types, we are developing and deploying a robust, yet flexible, data management and analysis platform that meets their immediate needs but is also extendable to the broader plant breeding community. In this poster, we present progress made over the past year working with partners at the iPlant Collaborative and the Breeding Management System software development team at CIMMYT, as well as plans for next steps in 2016.

### Genomes to Fields (G2F) Initiative

**Objective** - Leverage genomic information with phenotypic and environmental data to enable working knowledge and prediction of plant performance under variable growing conditions.

### GxE Sub-project

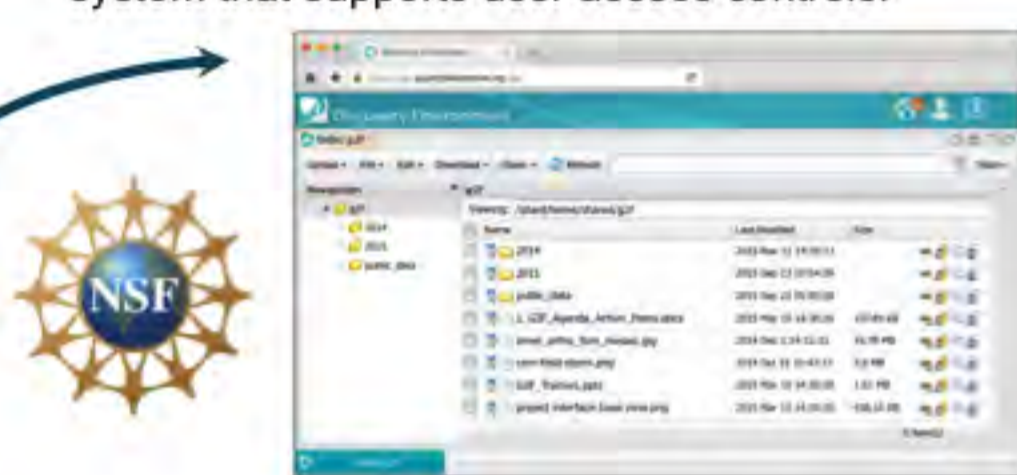
**Objective** - Evaluate environmental effects on a diverse collection of maize germplasm.

### Data Management Platform Components: iPlant, BMS, and GOBii

The iPlant Collaborative website features several key components:

- Discovery Environment:** Use hundreds of bioinformatics Apps and manage data in a single web interface.
- Atmosphere:** Create a custom cloud-based scientific analysis platform or use a ready-made one for your area of scientific interest.
- DNA Subway:** Take DNA Subway to teach classroom-friendly bioinformatics for genome analysis, DNA Barcoding, and RNA-Seq.
- Data Store:** Store, manage, access, and share all the data related to your research.
- Bisque Image Analysis Environment:** Exchange, explore, and analyze biological images and their metadata.

The **iPlant Discovery Environment** is a cloud based file storage system that supports user access controls.



Atmosphere supports a permanent CentOS virtual server where BMS has been re-coded to support web access.

**Integrated Breeding Platform (BMS)**  
Today's tools for tomorrow's crops.

BMS is a web accessible data store of:

- traits
- pedigrees
- phenotypes
- germplasm
- seed inventories



The iPlant-powered **Wiki** provides the community a place to store SOP's and documentation. User access controls assist with data integrity and embedded Google Sheets provides cross-platform/device data access.

The embedded meta-data worksheets are accessible through the web with out log-in.

Conditional formatting aid with data integrity and GPS coordinates are verified on Google maps.



**GOBii** is designed to store and access high-density genotypic data in condensed format, interoperates with BMS pedigrees for alleles, haplotypes, IBD blocks, etc.

### Phenotype Data Collected

Trait	Abbrev	Unit	Timing	Description/Procedure	Hybrid/Inbred	Notes/Considerations and Comments
Stand Count	SC	count	at harvest	Number of plants per plot at harvest.	*	Main consideration is how many plants were in the plot at harvest time. Counting can occur earlier but if a plot damage occurs before harvest they will need to be re-counted.
Plant Weight	PLW	gram [g]	at harvest	Shelled grain weight per plot	*	
Grain Moisture	SMI	percent [%]	at harvest	Moisture content in grain at harvest.	*	
Test Weight	TWT	kg/hL	at harvest		*	
Stalk Lodging	SLD	count	before harvest	Number of plants broken between ground level and top ear node at harvest.	*	Emphasis is on the number of plants, not the % which does not tell us much. Accurate stand counts and lodging counts are essential and will be used to calculate a % lodging in later analysis.
Green Snap (Regional)	GSP	count and date of harvest event	before flowering	Number of plants broken between ground level and top ear node before flowering.	*	
Stalk Lodging	SLD	count	before harvest	Number of plants that show root lodging per plot. I.e. those stems that lean substantially to one side (> 20% from vertical). Count includes "spooned" plants that have "straightened up" after becoming lodged earlier in the season.	*	Emphasis is on the number of plants, not the % which does not tell us much. Accurate stand counts and lodging counts are essential and will be used to calculate a % lodging in later analysis.
Archival	CHM	date [DD/MM/YYYY]		Days between planting and 50% of plants of a plot which neither ear or root on more than half of the main harvest cycle.	*	
Stilling	STP	date [DD/MM/YYYY]		Days between planting and 50% of plants of a plot show 10% emergence.	*	
Ear Height	EH	centimeter [cm] and date measured [DD/MM/YYYY]	after flowering	Placing measuring stick on ground next to the ear rows, "ear height" is measured at the primary ear-bearing node.	*	One plant is considered sufficient since these are inbreds and hybrids and are not segregating for traits.
Plant Height	PH	centimeter [cm] and date measured [DD/MM/YYYY]	after flowering	Measure the distance between the base of a plant and the top of the flag leaf.	*	One plant is considered sufficient since these are inbreds and hybrids and are not segregating for traits.
Ear Length	EL	centimeter [cm]		Length of ear	*	Note: This trait will be measured and recorded via iPlant. This trait will be measured and recorded via iPlant. This trait will be measured and recorded via iPlant.
Ear Width	EW	centimeter [cm]		Width of ear	*	Note: This trait will be measured and recorded via iPlant. This trait will be measured and recorded via iPlant. This trait will be measured and recorded via iPlant.
Harvest Row Spacing	RRS	count		Number of harvest rows. RRS is taken as the number of harvest rows around the middle to lower third of the	*	Note: This trait will be measured and recorded via iPlant. This trait will be measured and recorded via iPlant. This trait will be measured and recorded via iPlant.

### 2016 Objectives

Develop additional data management resources:

- Improved deployment of the BMS
- pipeline for GxE trial and nursery management
- tools for capturing environmental data
- enhanced seed inventory tracking
- tiered levels of BMS access for GxE cooperators
- Work with iPlant to develop:
  - a dedicated life sciences project management interface
  - an image analysis pipeline using the BISQUE

### Consortium Participants

#### G2F Collaborators

- Martin Bohn (UIUC)\*
- Ed Buckler (ARS)
- Darwin Campbell (ISU)\*
- James Clohessy (Cornell)
- Michael Coen (UW)
- Carolyn Lawrence Dill (ISU)\*
- Liang Dong (ISU)
- Jode Edwards (ARS)\*
- David Ertl (IA Corn)\*
- Sherry Flint-Garcia (ARS)\*
- Joseph Gage (UW)
- Jack Gardiner (ISU)\*
- Byron Good (Guelph)
- Mike Gore (Cornell)
- Patricio Grassini (UNL)
- Jerry Hatfield (ARS)
- Diego Jarquin (UNL)\*
- Shawn Kaeppler (UW)
- Liz Lee (Guelph)\*
- Natalia de Leon (UW)\*
- Zhizhai Liu (TAMU)
- Aaron Lorenz (UMN)
- Jonathan Lynch (PSU)
- Nathan Miller (UW)
- Jane Petzoldt (UW)\*
- Seth Murray (TAMU)\*
- Cinta Romay (Cornell)
- James Schnable (UNL)
- Pat Schnable (ISU)
- Nathan Springer (UMN)
- Edgar Spalding (UW)
- Srikant Srinivasan
- Kelly Thorp (ARS)
- Rod Williamson (IA Corn)
- Randy Wisser (Udel)
- Jianming Yu (ISU)

#### GxE Cooperators

- Martin Bohn (UIUC)\*
- Ed Buckler (ARS)
- Ignacio Ciampitti (KSU)
- Jode Edwards (ARS)\*
- Sherry Flint-Garcia (ARS)\*
- Christopher Graham (SDSU)
- Candy Hirsch (UMN)
- Jim Holland (ARS)
- David Hooker (Guelph)
- Shawn Kaeppler (UW)\*
- Joe Knoll (ARS)
- Judith Kolkman (Cornell)
- Greg Kruger (UNL)
- Nick Lauter (ARS)
- Aaron Lorenz (UMN)
- Liz Lee (Guelph)\*
- Natalia de Leon (UW)\*
- Jonathan Lynch (PSU)
- Steve Moose (UIUC)
- Seth Murray (TAMU)\*
- Rebecca Nelson (Cornell)
- Torbert Rocheford (Purdue)
- Oscar Rodriguez (UNL)
- Cinta Romay (Cornell)
- James Schnable (ISU)
- Brian Scully (ARS)
- Margaret Smith (Cornell)
- Nathan Springer (UMN)
- Peter Thomison (OSU)
- Mitch Tuinstra (Purdue)
- Randy Wisser (Udel)
- Wenwei Xu (TAMU)

\* GxE Data Management Team  
◇ G2F co-lead  
G2F Executive Committee



### Consortium Sponsors

### Data Platform Development Support