



# Phenome2020 - G2F Workshop

Alejandro Castro Aviles

Tucson, AZ

February 24<sup>th</sup>, 2020

[www.Genomes2Fields.org](http://www.Genomes2Fields.org)

# WELCOME!

- ❖ **Participants:**
  - ❖ 37 national
  - ❖ 23 international
  - ❖ 53 public sector
  - ❖ 7 private sector

# AGENDA:

- 11:15 – 11:25AM – Welcome and Introductions
- 11:25 – 12: 10PM – G2F Updates presentations
- 12:10 – 12:40PM – *Break and executive lunch served*
- 12:40 – 2:20PM – Research Presentations – first session
- 2:20 – 2:45PM – *Coffee break*
- 2:45 – 4:00PM – Research Presentations – second session
- 4:00PM – *Adjourn*

# ACKNOWLEDGMENTS:

- ❖ Teresa Myers – ASPB Meetings Coordinator
- ❖ All G2F Cooperators and Collaborators
- ❖ All Presenters and Participants
- ❖ G2F Executive Committee



# Genomes To Fields Collaborators and Cooperators

- Tim Beissinger (Göttingen)
- Martin Bohn (UIUC)
- Ed Buckler (ARS)
- Darwin Campbell (ISU)
- Alejandro Castro (UW)
- Ignacio Ciampitti (KSU)
- Liang Dong (ISU)
- Jode Edwards (ARS)
- David Ertl (IA Corn)
- Sherry Flint-Garcia (ARS)
- Christopher Graham (SDSU)
- Candy Hirsch (UMN)
- Jim Holland (ARS)
- Elizabeth Hood (AR State)
- David Hooker (Guelph)
- Joseph Gage (Cornell)
- Jack Gardiner (ISU)
- Fiona Goggin (AR State)
- Byron Good (Guelph)
- Mike Gore (Cornell)
- Patricio Grassini (UNL)
- Jerry Hatfield (ARS)
- Diego Jarquin (UNL)
- Shawn Kaeppler (UW)
- Joe Knoll (ARS)
- Greg Kruger (UNL)
- Nick Lauter (ARS)
- Carolyn Lawrence-Dill (ISU)
- Liz Lee (Guelph)
- Natalia de Leon (UW)
- Sanzchen Liu (Kansas)
- Argelia Lorence (AR State)
- Aaron Lorenz (UMN)
- Jonathan Lynch (PSU)
- Bridget McFarland (UW)
- John McKay (Colorado)
- Nathan Miller (UW)
- Steve Moose (UIUC)
- Seth Murray (TAMU)
- Rebecca Nelson (Cornell)
- Torbert Rocheford (Purdue)
- Oscar Rodriguez (UNL)
- Cinta Romay (Cornell)
- James Schnable (UNL)
- Pat Schnable (ISU)
- Brian Scully (ARS)
- Rajandeep Sekhon (Clemson)
- Kevin Silverstein (UMN)
- Maninder Singh (MI State)
- Margaret Smith (Cornell)
- Edgar Spalding (UW)
- Nathan Springer (UMN)
- Srikant Srinivasan (ISU)
- Yiwei Sun (ISU)
- Kelly Thorp (ARS)
- Kurt Thelen (MSU)
- Peter Thomison (OSU)
- Addie Thompson (MI State)
- Mitch Tuinstra (Purdue)
- Jason Wallace (UGA)
- Rod Williamson (IA Corn)
- Randy Wisser (UDeI)
- Wenwei Xu (TAMU)
- Jianming Yu (ISU)



# Genomes To Fields Sponsors





Genomes to Fields (G2F) Initiative Description  
and 2019 Data Summary

Alejandro Castro, G2F Collaborators

Phenome 2020 Conference

February 24

# the Genomes to Fields initiative

- ❖ A publicly led research initiative to catalyze and coordinate research linking genomics and predictive phenomics to achieve advances that generate economical and societal benefits
- ❖ Overall goal is to create crop varieties that are more resilient and better adapted to current and future variable environments
- ❖ We can achieve this by developing knowledge, tools and technologies that improve phenomic predictability and facilitate deployment of tools and resources that help address fundamental problems

# G2F – Executive Committee:



## G2F Executive Committee



Patrick Schnable **Co-Lead**  
Iowa State University



Natalia de Leon **Co-Lead**  
University of Wisconsin



Jonathan Lynch  
Pennsylvania State University



Nathan Springer  
University of Minnesota



Ed Buckler  
USDA-ARS



Shawn Kaeppler  
University of Wisconsin



David Ertl  
Iowa Corn Growers' Association

- Advocates for federal funding for initiative in area of predictive phenomics
  - Modeling, analytical, and HT phenotyping tools
  - Database development
  - Student training
- As part of the G2F initiative is the **Maize GxE Project**



Alejandro Castro,  
Research Coordinator

G2F  
Overview

# Overview of Maize GXE Project:

- G2F has been able to successfully evaluate a diverse set of hybrids across multiple locations and years to maximize interaction between representative set of material and multiple environments
- Living infrastructure of researchers across disciplines and a network of field phenotyping sites

	2014	2015	2016	2017	2018	2019
No. of unique locations	19	24	34	38	31	35
No. of plots planted	12,679	17,350	16,480	21,252	27,298	27,291
No. of plots post-harvest and data cleaning	12,302	11,956	16,014	15,940	24,379	TBD

## Since 2014:

- 182 unique environments
- ~122,350 plots
- ~ 12,000 unique varieties (with genetic sequence)

## 2019 Principal Investigators

### Arkansas (AR)

Beth Hood

### Colorado (CO)

John McKay

### Delaware (DE)

Randy Wisser

### Georgia (GA)

Joe Knoll  
Jason Wallace

### Illinois (IL)

Martin Bohn

### Indiana (IN)

Mitch Tunistra

### Iowa (IA)

Jode Edwards  
Patrick Schnable

### Kansas (KS)

Sanzhen Liu

### Michigan (MI)

Maninder Singh  
Addie Thompson

### Minnesota (MN)

Candice Hirsch  
Nathan Springer

### Missouri (MO)

Sherry Flint-Garcia

### Nebraska (NE)

James Schnable

### New York (NY)

Ed Buckler  
Michael Gore  
Rebecca Nelson

### North Carolina (NC)

Jim Holland

### Ohio (OH)

Richard Minyo  
Peter Thomison

### South Carolina (SC)

Rajan Sekhon

### Texas (TX)

Seth Murray  
Wenwei Xu

### Wisconsin (WI)

Natalia de Leon  
Shawn Kaepler

### Ontario (ON) - Canada

David Hooker  
Elizabeth Lee

### Other Investigators

(Not shown on map)

Tim Beissinger (Germany)

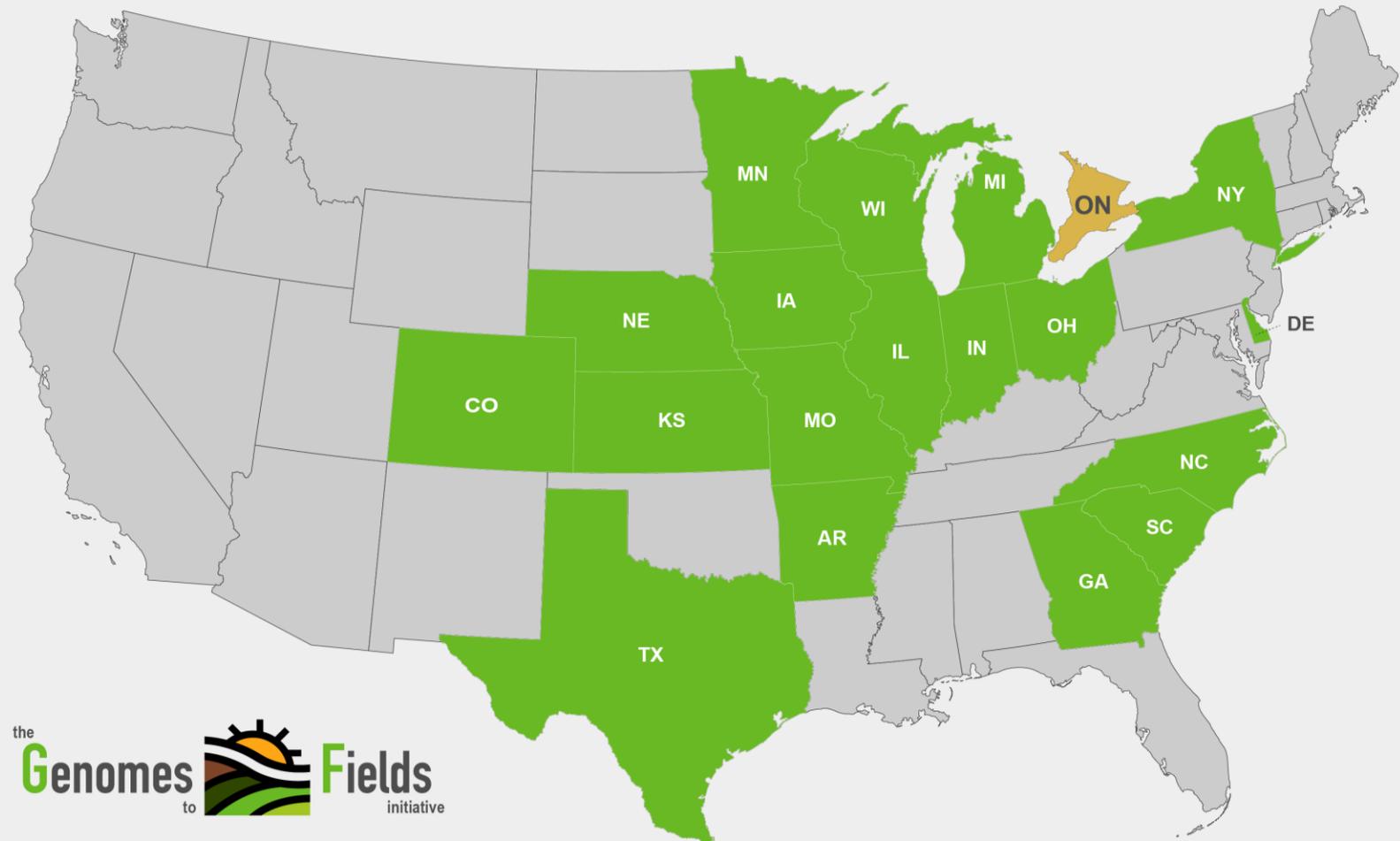
Arkansas State University  
Clemson University  
Colorado State University  
Cornell University  
Iowa State University  
Kansas State University

Michigan State University  
North Carolina State University  
Ohio State University  
Purdue University  
Texas A&M University  
University of Delaware

University of Georgia  
University of Guelph  
University of Illinois  
University of Minnesota  
University of Missouri  
University of Nebraska

University of Wisconsin  
USDA-ARS  
Georg-August-Universität Göttingen  
(Not shown on map)

## 2019 Academic & Federal Institutions



2019 GxE Principal Investigators and Academic & Federal Institutions in the United States and Canada

# Maize GXE Project - Field Design:

- Plots per location
  - Each location grows between 500 – 1,500 of various hybrids
  - Mega-location at least 1,500 plots and multiple testers
  - Incomplete blocks per location, overall balanced
  - 2 row plots
- Populations and testers change every two years
  - 2014-2015
  - 2016-2017
  - 2018-2019
- Yellow Stripe (YS) Hybrids
  - Common set of ~30 hybrids replicated in each location
  - Serve as checks across multiple locations and years
  - Ties data together
  - Used for specific research projects

Yellow Stripes
2369/LH123HT
B14A/H95
B14A/MO17
B14A/OH43
B37/H95
B37/MO17
B37/OH43
B73/MO17
B73/PHM49
B73/PHN82
CG119/CG108
CG44/CGR01
F42/H95
F42/MO17
F42/OH43
LH74/PHN82
PHG39/PHN82
PHW52/PHM49
PHW52/PHN82
TX110/87916
TX714/TX779
TX777/LH195
TX779/LH195
B73/TX779
LH132/TX779

# Data Collection

- Phenotypic:
  - Plot-level notes of stand count, lodging, flowering, yield, etc. ( $n=15$ )

## Agronomic Characteristics



## Yield Components



## Flowering/Silking Dates



## Plant/Ear Height



# Data Collection



Cinta Romay

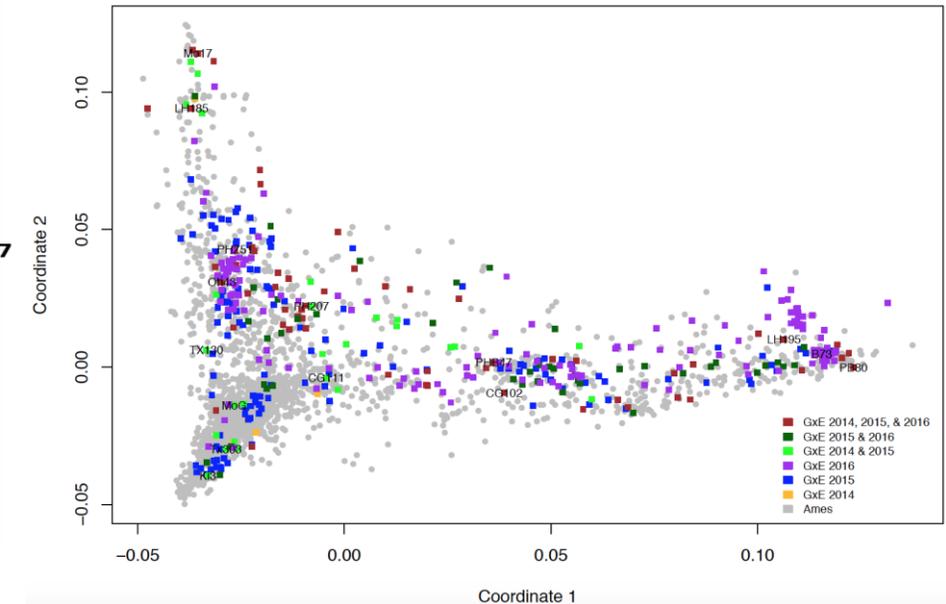
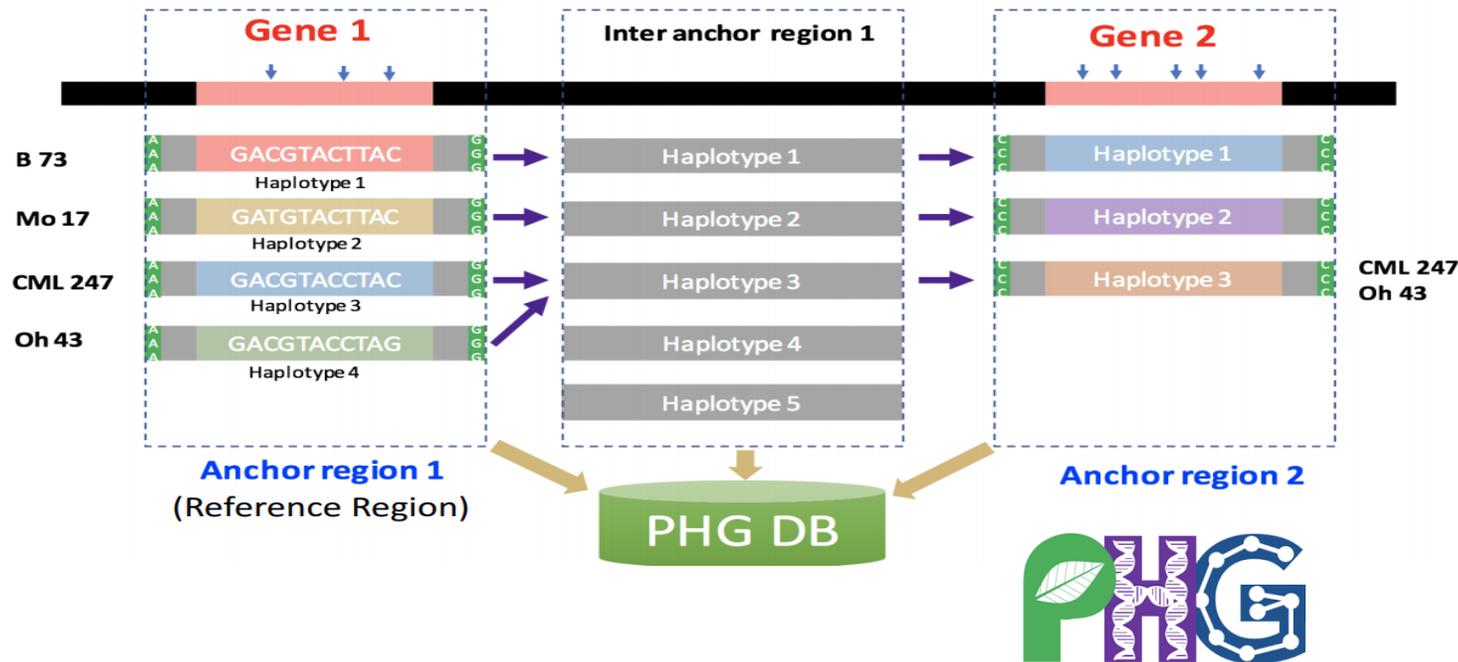


Ed Buckler

- Genotypic:
  - 2014 – 2017 Genotyping by Sequencing (GBS)
  - 2018 – 2019 Practical Haplotype Graph (PHG), focus on inter-genic region



Cornell University

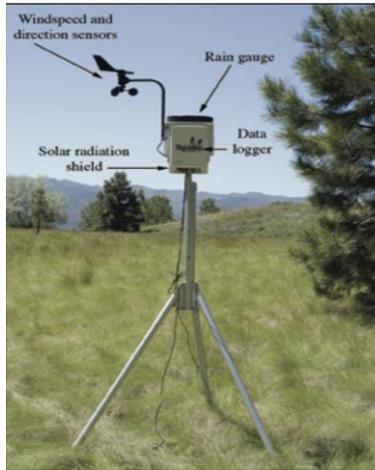


# Data Collection

## Weather Data:

- Field-level from weather stations temperature, relative humidity, rainfall, wind, PAR, soil temp, photoperiod, etc. ( $n=14$ )

Main Weather Station



MicroStation



## Soil Data:

- Soil Texture
- Soil Composition: %Sand, Silt, and Clay
- N, K, Zn, Fe, Mn, Cu, Ca, Na levels
- Soil pH



## MetaData:

- Field applications
- field GPS coordinates,
- specific machinery,
- field layout
- Any issues throughout the season ( $n=30+$ )

Experiment Code -->	WIH1	
When all values have been recorded, choose "done" --> (Data will not be collected until then)	Done	-
<b>Experiment</b>	<b>Treatment</b>	<b>Response</b>
	Standard	-
	City	Madison
	Farm	West Madison Agricultural Research Station
	Field	M1400 N
	Trial ID (Assigned by collaborator for internal reference)	
	Soil taxonomic ID and horizon description, if known	
<b>Weather station information</b>		
WatchDog 2700 weather station		
Weather station serial number (Last four digits, e.g. m2700a####)	8648	
Location		Not Sure?
Weather station location latitude (in decimal numbers NOT DMS)	43.057176	Check it!
Weather station location longitude (in decimal numbers NOT DMS)	-89.531052	
Date weather station placed in service	5/2/2017	
Date weather station removed from service	10/18/2017	
In-field pup weather station (if applicable)		
In-field station serial number	14820	
Location		Not Sure?
In-field station latitude (in decimal numbers NOT DMS)	43.056986	Check it!
In-field station longitude (in decimal numbers NOT DMS)	-89.530831	
In-field pup weather station (if applicable)		
In-field station serial number		
Location		Not Sure?
In-field station latitude (in decimal numbers NOT DMS)	43.07471	Check it!
In-field station longitude (in decimal numbers NOT DMS)	-89.384397	
Additional weather data source (if necessary)		
Link to additional weather source available online		
<b>Plot layout and planter information</b>		
Previous Crop	soybeans	
Pre-plant tillage methods)	field cultivates	
In-season tillage methods)	hand weed	
Plot length (center-to-center alley in feet)	10	Centimeter to Inch converter
Alley length (in inches)	30	Centimeters =
Row spacing (in inches)	30	
Types of planter (fluted cone, bell cone, air planter)	fluted cone	Inches =
Number kernels planted per plot (>200 seed/pack for cone planters)	80	

# Data Collection – SOP:

- Collaborators are provided a SOP and Phenotype Handbook each year (genomes2fields.org)

## Document Contents

I.	GxE Field Experiment 2019 Checklist	3
II.	Field Layout for 2018 and 2019 Seasons	4
III.	WatchDog 2700 Weather Station Configuration	6
IV.	Seed Information	6
V.	Field Metadata Collection	6
VI.	Phenotype and Performance Data Collection	7
VII.	Uploading Data to CyVerse	8
	Appendix A: Phenotyping Handbook	9
	Appendix B: Soil Sampling Handbook	15
	Appendix C: Additional Weather Station Resources	16

# Data Collection – SOP:

## Ear Height (EHT)

### Description/Procedure:

Placing measuring stick on ground next to the root crown, “ear height” is measured at the primary ear bearing node. See *Picture 1*.

**Timing:** At plant maturity

**n =** 1 representative plant per plot

**Unit:** centimeter [cm]

**Notes:** One plant is considered sufficient since these are inbreds and hybrids and are not segregating for traits.



*Picture 1*



# Data Cleaning – Data Wrangling:

- Group of PIs, data scientists, post docs and grad students
- Eliminates obvious outliers, no bias
- Examples of cleaning steps:
  - Stand count > # of seeds planted, then “NA”
  - Ear height > plant height, then “NA”
  - Plot weight, test weight, grain moisture < 0 or > 400, then “NA”
  - Grain yield > 500, then “NA”
  - Spatial variability using field maps for high/low values
  - Read through metadata to see if additional plots need to be discarded

G2F  
Overview

Data  
Collection

Data Cleaning

# Weather Data Cleaning:

- Impute missing weather data from:
  - Nearest airport
  - Publicly available weather stations

The screenshot shows the Iowa State University Iowa Environmental Mesonet website. The header includes the university name and logo, a search bar, and navigation links for CONTACT US, DISCLAIMER, and APPS. Below the header is a navigation menu with options like Archive, Climate, Current, Info, GIS, Networks, Roads, Svr Wx, and Webcams. The main content area is titled "ASOS Network" and "ASOS-AWOS-METAR Data Download".

The IEM maintains an ever growing archive of automated airport weather observations from around the world! These observations are typically called 'ASOS' or sometimes 'AWOS' sensors. A more generic term may be METAR data, which is a term that describes the format the data is transmitted as. If you don't get data for a request, please feel free to [contact us](#) for help. The IEM also has a one minute interval dataset for Iowa ASOS (2000-) and AWOS (1995-2011) sites. This archive simply provides the as-is collection of historical observations, very little quality control is done. "M" is used to denote missing data.

Here is a [python script example](#) that automates the download of data from this interface. A community user has contributed [R language](#) version of the python script.

Please be patient with this page as it will take a number of seconds to process your request and provide the result.

Select Network:

1) Select Station/Network by clicking on location:

Select Widget for NC\_ASOS Network

Sort Available Stations:  Selected Stations:

[1A5] FRANKLIN  
[24A] Sylva  
[2DP] Dare County Gunny Range  
[5W8] Silver City

Enter some text here to f

2) Select From Available Data:

All Available  
Air Temperature [F]  
Air Temperature [C]  
Dew Point [F]  
Dew Point [C]  
Relative Humidity [%]  
Heat Index/Wind Chill [F]  
Wind Direction  
Wind Speed [knots]  
Wind Speed [mph]

3) Specific Date Range (If needed):

**Note:** There is up to an hour delay for observations to appear within this service. A process runs at approximately 20 minutes after the hour to copy over the previous hour's worth of data in the backend database behind this service.

Start Date:     
End Date:

4) Timezone of Observation Times:

The following options are available for how the observation time is presented.

5) Download Options:

Data Format:   
Include Latitude + Longitude?   
How to represent missing data?   
How to represent Trace reports?

G2F  
Overview

Data  
Collection

Data Cleaning



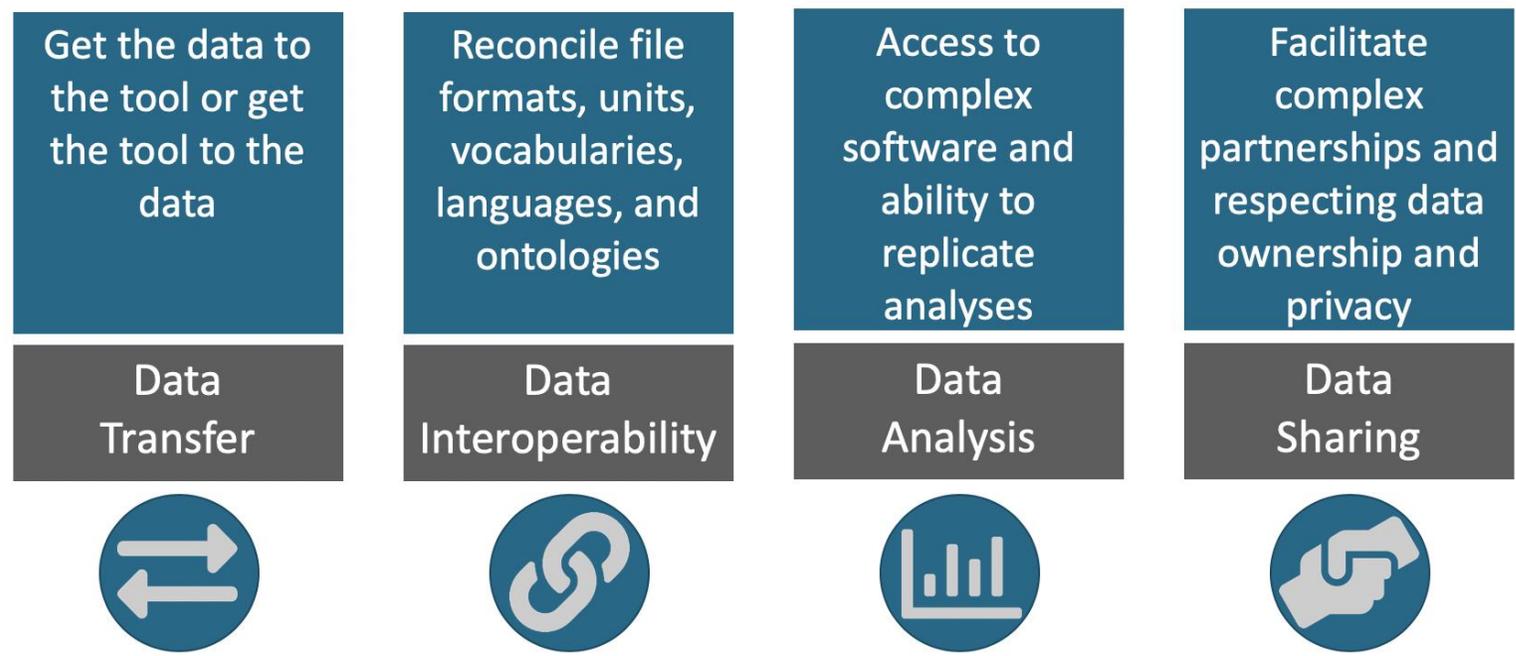
GEMS

Data-Driven Agricultural Innovation  
GENETICS • ENVIRONMENT • MANAGEMENT • SOCIOECONOMICS

# Data Processing:



- GEMS is a novel data sharing and big data analytical platform;
- Kevin Silverstein, Bioinformaticist/GEMS Operations Manager



# Data Is Made Publicly Available:

[Home](#)[About](#)[Funded Projects](#)[Publications](#)[Meetings](#)[Resources](#)

## Data Downloads

Field Season	Release to Collaborators (ARK)	Release to Public (DOI)
2014	Not Available	<a href="https://doi.org/10.7946/P2V888">https://doi.org/10.7946/P2V888</a>
2015	February 2017	<a href="https://doi.org/10.7946/P24S31">https://doi.org/10.7946/P24S31</a>
2016	February 2017	<a href="https://doi.org/10.7946/P3MW27">https://doi.org/10.7946/P3MW27</a>
2017	February 2018	March 2019
N	February of N+1	March of N+2

ⓘ Please report data errors or inconsistencies that you might identify [using this submission form](#)

**2016 Field Season (Available to Public):** Phenotypic, genotypic, and environment data for the 2016 field season: <https://doi.org/10.7946/P3MW27>

**2015 Field Season (Available to Public):** Phenotypic, genotypic, and environment data for the 2015 field season: <https://doi.org/10.7946/P24S31>

**2014 Field Season (Available to Public):** Phenotypic, genotypic, and environment data for the 2014 field season: <https://doi.org/10.7946/P2V888>

[›Ear Imaging Data](#)[›Data Downloads](#)[›Standard Operating Procedures \(SOP\)](#)[›Newsletters](#)

- › 2018-12-31
- › 2018-07-18
- › 2017-08-11
- › 2017-04-10
- › 2016-06-21

[›Genomes2Fields Links](#)[›Genomes2Fields Logos & Images](#)[›Photos & Videos](#)

DATA NOTE

Open Access



## Maize Genomes to Fields: 2014 and 2015 field season genotype, phenotype, environment, and inbred ear image datasets

Naser AlKhalifah<sup>1,23†</sup>, Darwin A. Campbell<sup>1†</sup>, Celeste M. Falcon<sup>2†</sup>, Jack M. Gardiner<sup>1,24†</sup>, Nathan D. Miller<sup>2†</sup>, Maria Cinta Romay<sup>3†</sup>, Ramona Walls<sup>4†</sup>, Renee Walton<sup>1†</sup>, Cheng-Ting Yeh<sup>1†</sup>, Martin Bohn<sup>5</sup>, Jessica Bubert<sup>5</sup>, Edward S. Buckler<sup>3,6</sup>, Ignacio Ciampitti<sup>7</sup>, Sherry Flint-Garcia<sup>6,8</sup>, Michael A. Gore<sup>3</sup>, Christopher Graham<sup>9</sup>, Candice Hirsch<sup>10</sup>, James B. Holland<sup>6,11</sup>, David Hooker<sup>12</sup>, Shawn Kaeppler<sup>2</sup>, Joseph Knoll<sup>6</sup>, Nick Lauter<sup>1,6</sup>, Elizabeth C. Lee<sup>13</sup>, Aaron Lorenz<sup>14,25</sup>, Jonathan P. Lynch<sup>15</sup>, Stephen P. Moose<sup>5</sup>, Seth C. Murray<sup>16</sup>, Rebecca Nelson<sup>3</sup>, Torbert Rocheford<sup>17</sup>, Oscar Rodriguez<sup>14</sup>, James C. Schnable<sup>14</sup>, Brian Scully<sup>6,18</sup>, Margaret Smith<sup>7</sup>, Nathan Springer<sup>10</sup>, Peter Thomison<sup>19</sup>, Mitchell Tuinstra<sup>17</sup>, Randall J. Wisser<sup>20</sup>, Wenwei Xu<sup>21</sup>, David Ertl<sup>22</sup>, Patrick S. Schnable<sup>1†</sup>, Natalia De Leon<sup>2†</sup>, Edgar P. Spalding<sup>2†</sup>, Jode Edwards<sup>1,6†</sup> and Carolyn J. Lawrence-Dill<sup>1†</sup>

### Abstract

**Objectives:** Crop improvement relies on analysis of phenotypic, genotypic, and environmental data. Given large, well-integrated, multi-year datasets, diverse queries can be made: Which lines perform best in hot, dry environments? Which alleles of specific genes are required for optimal performance in each environment? Such datasets also can be leveraged to predict cultivar performance, even in uncharacterized environments. The maize Genomes to Fields (G2F) Initiative is a multi-institutional organization of scientists working to generate and analyze such datasets from existing, publicly available inbred lines and hybrids. G2F's genotype by environment project has released 2014 and 2015 datasets to the public, with 2016 and 2017 collected and soon to be made available.

**Data description:** Datasets include DNA sequences; traditional phenotype descriptions, as well as detailed ear, cob, and kernel phenotypes quantified by image analysis; weather station measurements; and soil characterizations by site. Data are released as comma separated value spreadsheets accompanied by extensive README text descriptions. For genotypic and phenotypic data, both raw data and a version with outliers removed are reported. For weather data, two versions are reported: a full dataset calibrated against nearby National Weather Service sites and a second calibrated set with outliers and apparent artifacts removed.

**Keywords:** Maize, Genome, Genotype, Environment, Breeding, Phenotype, Prediction, Soil, Inbred, Hybrid

\*Correspondence: dertl@iowacorn.org; schnable@iastate.edu; ndeleongatti@wisc.edu; spalding@wisc.edu; jode.edwards@ars.usda.gov; trfd@iastate.edu

<sup>†</sup>Naser AlKhalifah, Darwin A. Campbell, Celeste M. Falcon, Jack M. Gardiner, Nathan D. Miller, Maria Cinta Romay, Ramona Walls, Renee Walton, Cheng-Ting Yeh are joint first authors

<sup>1</sup>Iowa State University, Ames, IA 50011, USA

<sup>2</sup>University of Wisconsin, Madison, WI 53706, USA

<sup>3</sup>USDA-ARS, Beltsville, MD, USA

<sup>4</sup>Iowa Corn Growers Association, Johnston, IA 50131, USA

Full list of author information is available at the end of the article



© The Author(s) 2018. This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated.

DATA NOTE

Open Access



## Maize genomes to fields (G2F): 2014–2017 field seasons: genotype, phenotype, climatic, soil, and inbred ear image datasets

Bridget A. McFarland<sup>1†</sup>, Naser AlKhalifah<sup>1†</sup>, Martin Bohn<sup>2</sup>, Jessica Bubert<sup>2</sup>, Edward S. Buckler<sup>3,4</sup>, Ignacio Ciampitti<sup>5</sup>, Jode Edwards<sup>4,6</sup>, David Ertl<sup>7</sup>, Joseph L. Gage<sup>3</sup>, Celeste M. Falcon<sup>1</sup>, Sherry Flint-Garcia<sup>4,8</sup>, Michael A. Gore<sup>3</sup>, Christopher Graham<sup>9</sup>, Candice N. Hirsch<sup>10</sup>, James B. Holland<sup>4,11</sup>, Elizabeth Hood<sup>12</sup>, David Hooker<sup>13</sup>, Diego Jarquin<sup>14</sup>, Shawn M. Kaeppler<sup>1</sup>, Joseph Knoll<sup>6</sup>, Greg Kruger<sup>14</sup>, Nick Lauter<sup>4,6</sup>, Elizabeth C. Lee<sup>13</sup>, Dayane C. Lima<sup>1</sup>, Aaron Lorenz<sup>10</sup>, Jonathan P. Lynch<sup>15</sup>, John McKay<sup>16</sup>, Nathan D. Miller<sup>1</sup>, Stephen P. Moose<sup>2</sup>, Seth C. Murray<sup>17</sup>, Rebecca Nelson<sup>3</sup>, Christina Poudyal<sup>10</sup>, Torbert Rocheford<sup>18</sup>, Oscar Rodriguez<sup>14</sup>, Maria Cinta Romay<sup>3</sup>, James C. Schnable<sup>14</sup>, Patrick S. Schnable<sup>6</sup>, Brian Scully<sup>19</sup>, Rajandeep Sekhon<sup>20</sup>, Kevin Silverstein<sup>10</sup>, Maninder Singh<sup>21</sup>, Margaret Smith<sup>3</sup>, Edgar P. Spalding<sup>1</sup>, Nathan Springer<sup>10</sup>, Kurt Thelen<sup>21</sup>, Peter Thomison<sup>22</sup>, Mitchell Tuinstra<sup>18</sup>, Jason Wallace<sup>23</sup>, Ramona Walls<sup>24</sup>, David Willis<sup>9</sup>, Randall J. Wisser<sup>25</sup>, Wenwei Xu<sup>17</sup>, Cheng-Ting Yeh<sup>6</sup> and Natalia de Leon<sup>1†</sup>

### Abstract

**Objectives:** Advanced tools and resources are needed to efficiently and sustainably produce food for an increasing world population in the context of variable environmental conditions. The maize genomes to fields (G2F) Initiative is a multi-institutional effort that seeks to approach this challenge by developing a flexible and distributed infrastructure addressing emerging problems. G2F has generated large-scale phenotypic, genotypic, and environmental datasets using publicly available inbred lines and hybrids evaluated through a network of collaborators that are part of the G2F's genotype-by-environment (G × E) project. This report covers the public release of datasets for 2014–2017.

**Data description:** Datasets include inbred genotypic information; phenotypic, climatic, and soil measurements and metadata information for each testing location across years. For a subset of inbreds in 2014 and 2015, yield component phenotypes were quantified by image analysis. Data released are accompanied by README descriptions. For genotypic and phenotypic data, both raw data and a version without outliers are reported. For climatic data, a version calibrated to the nearest airport weather station and a version without outliers are reported. The 2014 and 2015 datasets are updated versions from the previously released files [1] while 2016 and 2017 datasets are newly available to the public.

**Keywords:** Maize, Genome, Genotype, GBS, G × E, Hybrid, Inbred, Phenotype, Environment, Field metadata

### Objective

Genomes to fields (G2F) is a multi-institutional, public collaborative to develop information and tools that support the translation of maize (*Zea mays* L.) genomic information into relevant phenotypes for the benefit of

\*Correspondence: ndeleongatti@wisc.edu

<sup>†</sup>Bridget A. McFarland and Naser AlKhalifah are joint first authors

<sup>1</sup>University of Wisconsin, Madison, WI 53706, USA

Full list of author information is available at the end of the article



© The Author(s) 2020. This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

# Published Research Involving G2F Data:

## Genomes To Fields (G2F) Publications

Symbol Legend: \* Authors contributed equally to the article



Field Crops Research. 2020, 249: 107737 | 1 April 2020  
doi:10.1016/j.fcr.2020.107737

### Stalk bending strength is strongly associated with maize stalk lodging incidence across multiple environments

Rajandeep S. Sekhon, Chase N. Joyner, Arlyn J. Ackerman, Christopher S. McMahan, Douglas D. Cook, Daniel J. Robertson

[Download Article](#) | [View Article](#)



BMC Research Notes. 2020, 13: 71 | 12 February 2020  
doi:10.1186/s13104-020-4922-8

### Maize genomes to fields (G2F): 2014–2017 field seasons: genotype, phenotype, climatic, soil, and inbred ear image datasets

Bridget A. McFarland\*, Naser AlKhalifah\*, Martin Bohn, Jessica Bubert, Edward S. Buckler, Ignacio Ciampitti, Jode Edwards, David Ertl, Joseph L. Gage, Celeste M. Falcon, Sherry Flint-Garcia, Michael A. Gore, Christopher Graham, Candice N. Hirsch, James B. Holland, Elizabeth Hood, David Hooker, Diego Jarquin, Shawn M. Kaeppler, Joseph Knoll, Greg Kruger, Nick Lauter, Elizabeth C. Lee, Dayane C. Lima, Aaron Lorenz, Jonathan P. Lynch, John McKay, Nathan D. Miller, Stephen P. Moose, Seth C. Murray, Rebecca Nelson, Christina Poudyal, Torbert Rocheford, Oscar Rodriguez, Maria Cinta Romay, James C. Schnable, Patrick S. Schnable, Brian Scully, Rajandeep Sekhon, Kevin Silverstein, Maninder Singh, Margaret Smith, Edgar P. Spalding, Nathan Springer, Kurt Thelen, Peter Thomison, Mitchell Tuinstra, Jason Wallace, Ramona Walls, David Wills, Randall J. Wisser, Wenwei Xu, Cheng-Ting Yeh, Natalia de Leon

[Download Article](#) | [View Article](#)



Frontiers in Plant Science. 2019, 10: 1550 | 12 December 2019  
doi:10.3389/fpls.2019.01550

### Millimeter-Level Plant Disease Detection From Aerial Photographs via Deep Learning and Crowdsourced Data

Tyr Wiesner-Hanks, Harvey Wu, Ethan Stewart, Chad DeChant, Nicholas Kaczmar, Hod Lipson, Michael A. Gore, Rebecca J. Nelson

[Download Article](#) | [View Article](#)



The Plant Genome Journal. 2019, 2(1): 190011 | 22 November 2019  
doi:10.2135/tppj2019.07.0011

### In-Field whole-plant maize architecture characterized by subcanopy rovers and latent space phenotyping

Joseph L. Gage, Elliot Richards, Nicholas Lepak, Nicholas Kaczmar, Chinmay Soman, Girish Chowdhary, Michael A. Gore, Edward S. Buckler

[Download Article](#) | [View Article](#)



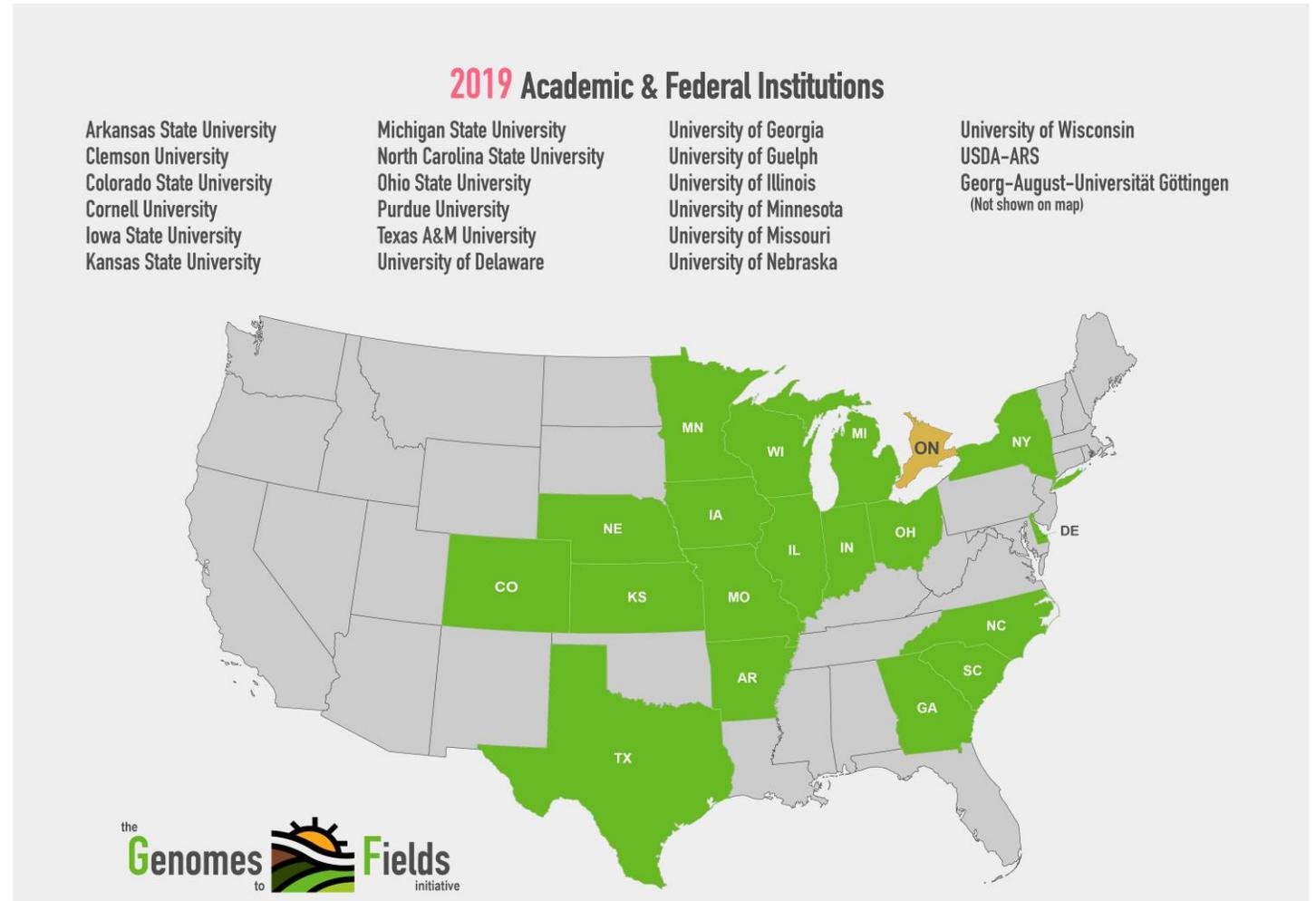
The Plant Phenome Journal. 2019, 2(1) | 7 November 2019  
doi:10.2135/tppj2019.03.0006

### Autonomous detection of plant disease symptoms directly from aerial imagery

Harvey Wu, Tyr Wiesner-Hanks, Ethan L. Stewart, Chad DeChant, Nicholas Kaczmar, Michael A. Gore, Rebecca J. Nelson, Hod Lipson

# Overview of 2019:

- 21 Academic and Federal Institutions
- 30 PIs
- 35 unique locations



G2F  
Overview

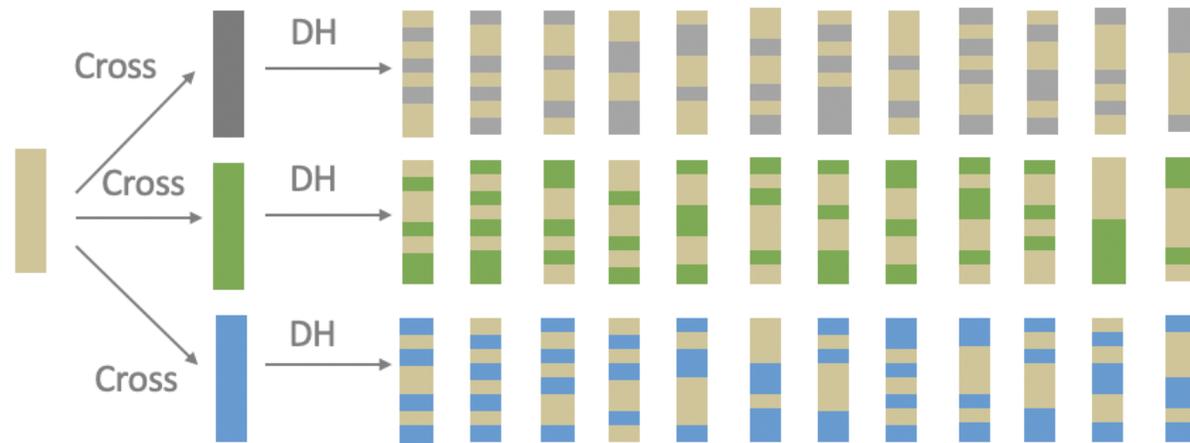
Data  
Collection

Data Cleaning

2019 Recap

# PHW65 mini-NAM Maize Population:

- PHW65 NAM-type population
  - PHW65 reference crossed with 3 lines contrasting levels of selection
  - DHs were derived from crosses
  - Location-specific testers were used to generate hybrids



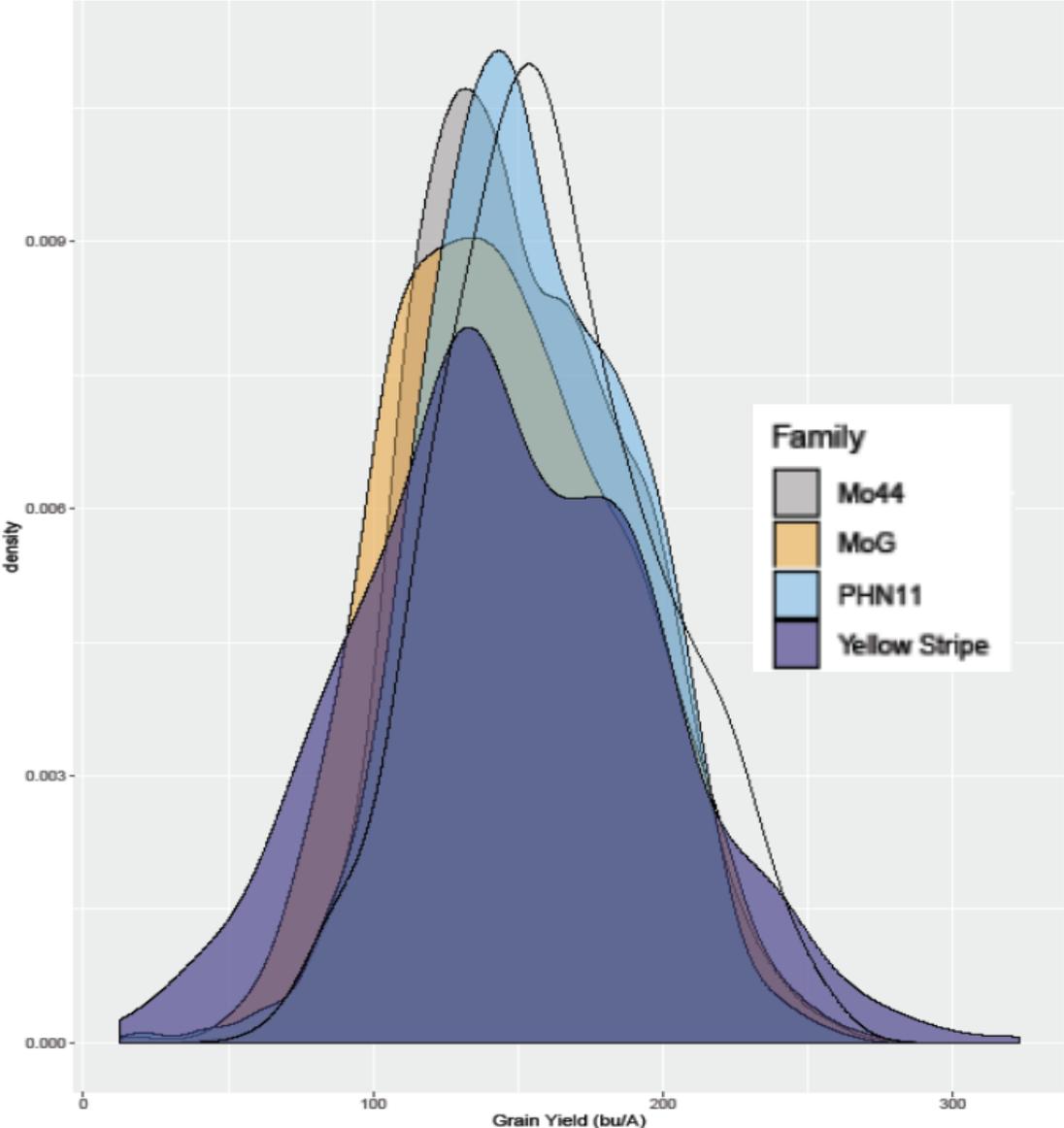
G2F  
Overview

Data  
Collection

Data Cleaning

2019 Recap

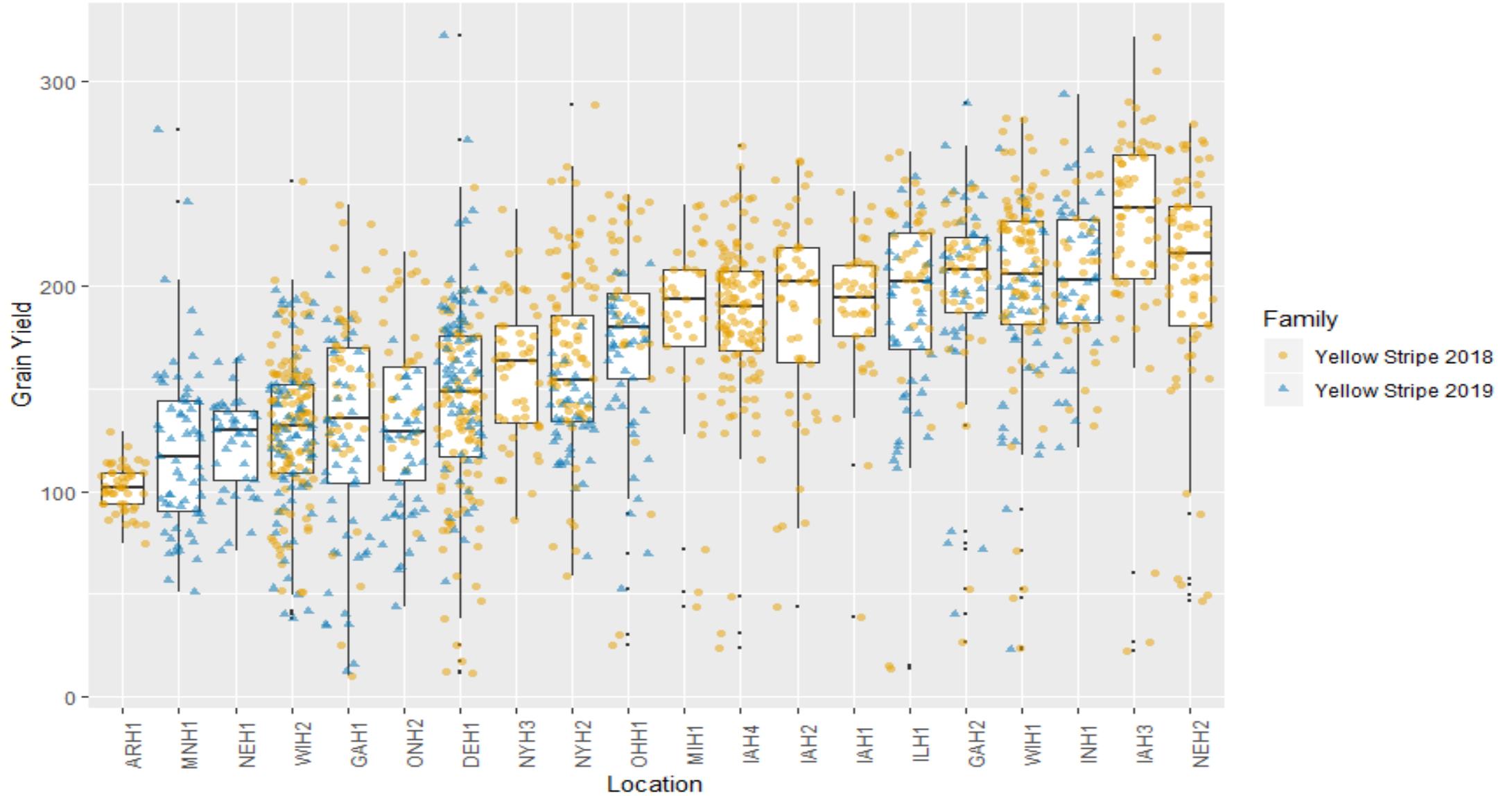
# Grain Yield – Per Family 2019:



# Grain Yield – Yellow Stripe:



# Grain Yield of Yellow Stripe.- Regression:



# Genomes To Fields Collaborators and Cooperators

- Tim Beissinger (Göttingen)
- Martin Bohn (UIUC)
- Ed Buckler (ARS)
- Darwin Campbell (ISU)
- Alejandro Castro (UW)
- Ignacio Ciampitti (KSU)
- Liang Dong (ISU)
- Jode Edwards (ARS)
- David Ertl (IA Corn)
- Sherry Flint-Garcia (ARS)
- Christopher Graham (SDSU)
- Candy Hirsch (UMN)
- Jim Holland (ARS)
- Elizabeth Hood (AR State)
- David Hooker (Guelph)
- Joseph Gage (Cornell)
- Jack Gardiner (ISU)
- Fiona Goggin (AR State)
- Byron Good (Guelph)
- Mike Gore (Cornell)
- Patricio Grassini (UNL)
- Jerry Hatfield (ARS)
- Diego Jarquin (UNL)
- Shawn Kaeppler (UW)
- Joe Knoll (ARS)
- Greg Kruger (UNL)
- Nick Lauter (ARS)
- Carolyn Lawrence-Dill (ISU)
- Liz Lee (Guelph)
- Natalia de Leon (UW)
- Sanzchen Liu (Kansas)
- Argelia Lorence (AR State)
- Aaron Lorenz (UMN)
- Jonathan Lynch (PSU)
- Bridget McFarland (UW)
- John McKay (Colorado)
- Nathan Miller (UW)
- Steve Moose (UIUC)
- Seth Murray (TAMU)
- Rebecca Nelson (Cornell)
- Torbert Rocheford (Purdue)
- Oscar Rodriguez (UNL)
- Cinta Romay (Cornell)
- James Schnable (UNL)
- Pat Schnable (ISU)
- Brian Scully (ARS)
- Rajandeep Sekhon (Clemson)
- Kevin Silverstein (UMN)
- Maninder Singh (MI State)
- Margaret Smith (Cornell)
- Edgar Spalding (UW)
- Nathan Springer (UMN)
- Srikant Srinivasan (ISU)
- Yiwei Sun (ISU)
- Kelly Thorp (ARS)
- Kurt Thelen (MSU)
- Peter Thomison (OSU)
- Addie Thompson (MI State)
- Mitch Tuinstra (Purdue)
- Jason Wallace (UGA)
- Rod Williamson (IA Corn)
- Randy Wisser (UDeI)
- Wenwei Xu (TAMU)
- Jianming Yu (ISU)



# Genomes To Fields Sponsors



# Questions?

Alejandro Castro

G2F Research Coordinator / University of Wisconsin -  
Madison

[castroaviles@wisc.edu](mailto:castroaviles@wisc.edu)

[genomes2fields.org](http://genomes2fields.org)

