

2020-2021 Field Seasons of Maize G x E Project within Maize Genomes to Fields Initiative

Dayane Cristina Lima (✉ dclima@wisc.edu)

University of Wisconsin – Madison

Alejandro Castro Aviles

University of Wisconsin – Madison

Ryan Timothy Alpers

University of Wisconsin – Madison

Alden Perkins

University of Wisconsin – Madison

Dylan L. Schoemaker

University of Wisconsin – Madison

Martin Costa

University of Wisconsin – Madison

Shawn Kaepler

University of Wisconsin – Madison

David Ertl

Iowa Corn Promotion Board

Maria Cinta Romay

Cornell University

Joseph L. Gage

North Carolina State University

James Holland

USDA-ARS Plant Science Research Unit

Timothy Beissinger

University of Göttingen, University of Göttingen Center for Integrated Breeding Research

Martin Bohn

University of Illinois at Urbana-Champaign

Edward Buckler

USDA-ARS and Cornell University

Jode Edwards

USDA ARS CICGRU

Sherry Flint-Garcia

USDA-ARS Plant Science Research Unit

Michael A. Gore

Cornell University

Candice N. Hirsch

University of Minnesota

Joseph E. Knoll

USDA-ARS Crop Genetics and Breeding Research Unit

John McKay

Colorado State University

Richard Minyo

The Ohio State University

Seth C. Murray

Texas A&M University

James Schnable

University of Nebraska–Lincoln

Rajandeep S. Sekhon

Clemson University

Maninder P. Singh

Michigan State University

Erin E. Sparks

University of Delaware

Peter Thomison

The Ohio State University

Addie Thompson

Michigan State University

Mitchell Tuinstra

Purdue University West Lafayette

Jason Wallace

University of Georgia

Jacob Washburn

USDA-ARS Plant Science Research Unit

Teclamarium Weldekidan

University of Delaware

Wenwei Xu

Texas A&M University

Natalia de Leon


University of Wisconsin – Madison

Data Note

Keywords: Maize, Genotype by Environment, Grain Yield, Stability, Prediction.

Posted Date: May 11th, 2023

DOI: <https://doi.org/10.21203/rs.3.rs-2908766/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Version of Record: A version of this preprint was published at BMC Research Notes on September 14th, 2023. See the published version at <https://doi.org/10.1186/s13104-023-06430-y>.

Abstract

Objectives:

This release note describes the Maize G x E project datasets within the Genomes to Fields (G2F) Initiative. The Maize G x E project aims to understand genotype by environment (G x E) interactions and use the information collected to improve resource allocation efficiency and increase genotype predictability and stability, particularly in scenarios of variable environmental patterns. Hybrids and inbreds are evaluated across multiple environments and phenotypic, genotypic, environmental, and metadata information are made publicly available.

Data description:

The datasets include phenotypic data of the hybrids and inbreds evaluated in 30 locations across the US and one location in Germany in 2020 and 2021, soil and climatic measurements and metadata information for all environments (combination of year and location), ReadMe, and description files for each data type. A set of common hybrids is present in each environment to connect with previous evaluations. Each environment had a collaborator responsible for collecting and submitting the data, the G x E coordination team combined all the collected information and removed obvious erroneous data. Collaborators received the combined data to use, verify and declare that the data generated in their own environments was accurate. Combined data is released to the public with minimal filtering to maintain fidelity to the original data.

Objective

The release of this data provides a unique resource to understand and dissect genotype-by-environment interactions in maize (*Zea mays* subsp. *mays* L.). Collaborators generate phenotypic, environmental, and metadata datasets to support a more comprehensive understanding of the opportunities and challenges associated with maize production in various environments. The Maize G x E project data is made available to the public in its original form, with minimum filtering to remove erroneous data or as specified by collaborators and in the description files. This approach ensures that the publicly available data contains the maximum amount of information collected by project collaborators and empowers users to define their quality controls based on their specific goals.

A set of 1184 publicly available hybrids were evaluated in the 2020 and 2021 seasons across 30 different locations. The main group of hybrids was produced by the cross of doubled-haploid (DH) inbred lines from the Wisconsin Stiff Stalk MAGIC population (WI-SS-MAGIC), crossed with three ex-PVP inbred testers, PHZ51, PHP02, and PHK76 [1]. The WI-SS-MAGIC population involves the inbreds B73, B84, NKH8431, LH145, PHB47, and PHJ40 as parents in the initial crosses, and a detailed description of the population creation and DH production is in Michel et al. (2022) [1]. The testers were selected to allow adaptation of materials to the wide array of maturities sampled across the project. Inbred tester PHZ51 was used in southern locations (DEH1, GAH1, GAH2, IAH1, IAH2, IAH3, IAH4, MOH1, NCH1, NEH1, NEH2, NEH3, NYH3, SCH1, TXH1, TXH2, TXH3, WIH2), PHK76 in the Midwest and intermediate locations (DEH1, IAH1, IAH2, IAH3, IAH4, ILH1, INH1, MOH1, NCH1, NYH3, WIH2), and PHP02 in the northern locations (DEH1, GEH1, IAH2, IAH3, IAH4, MIH1,

MNH1, MOH1, NYH2, OHH1, WIH1, WIH2, WIH3). Five locations (mega-locations) had hybrids created using all three testers (DEH1, IAH3, IAH4, MOH1, WIH2). Additional smaller-scale experiments were conducted alongside the main experiment for additional phenotyping and/or deployment of novel phenotyping methods and tools across approximately 83% of the locations. These experiments included the external Yellow Stripes (YS), same set of check hybrids know as 'Yellow Stripe' used to connect location and years, but evaluated in a different experiment; the High-Intensity Phenotyping Site (HIPS), which tested 22 hybrids (HIP_Hybrid) and 22 inbreds (HIP_Inbred). HIPS was introduced in 2020 as a more comprehensive phenotyping set for aerial high-throughput phenotyping platforms. The choice of hybrids and inbreds was based on their historical importance and relevance to other connected projects.

Data description

The 2020 and 2021 datasets are publicly available via CyVerse/iPlant. These datasets contain phenotypic, environmental, soil, and supplemental data, and have been structured according to the specifications outlined in Table 1.

- **Phenotypic data:** Phenotypic measurements that follow a standard set of instructions, available at genomes2fields.org. Standard traits include days to anthesis, days to silking, ear height, plant height, stand count, stalk lodging, root lodging, grain moisture, test weight, plot weight, and grain yield. Both raw data and minimally quality-controlled (clean) data are reported separately. Out of range observations were set to missing values following the rules described in the readMe files.
- **Environmental dataset:** WatchDog 2700 weather stations (Spectrum Technologies) were placed at each field site. Data were collected at 30-min intervals, or according with collaborator set up, from planting through harvest at each location. The geographic locations of the experiments are not identical across years due to crop rotation management practices; thus, the locations of the weather stations vary across years. Each station measured wind speed, direction, and gust; air temperature, dewpoint, relative humidity; soil temperature and moisture; rainfall and solar radiation. Additional measurements taken at selected sites included soil electrical conductivity, ultra-violet light, carbon dioxide, and photosynthetically active radiation. Instructions for weather station maintenance activities including pre-season tasks, field setup, maintenance throughout the growing season, and removal are available on the G2F webpage [2].
- **Soil dataset:** Each field location collected soil samples that represent the experiment field according to the instructions available on the G2F webpage.
- **Supplemental dataset:** Supplemental information consists of metadata (any field-level data collected at planting, in season, and/or at harvest), agronomic information (list of products, nutrients, and irrigation applied), and cooperator list (collaborators responsible for the field locations in 2020 and 2021).

Table 1: Overview of 2020 and 2021 planting seasons datasets.

Label	Name of data file/data set	File types (Extension)	Data repository and identifier
2020 Planting Season	readme.txt	.txt	CyVerse [3]
	a._2020_phenotypic_data	directory	(https://doi.org/10.25739/hzzs-a865)
	g2f_2020_phenotypic_data_description.pdf	.pdf	
	g2f_2020_phenotypic_raw_data.csv	.csv	
	g2f_2020_phenotypic_clean_data.csv	.csv	
	b._2020_weather_data	directory	
	g2f_2020_weather_data_description.pdf	.pdf	
	g2f_2020_weather_readMe.txt	.txt	
	2020_weather_raw.csv	.csv	
	2020_weather_cleaned.csv	.csv	
	c._2020_soil_data	directory	
	g2f_2020_soil_data_description.pdf	.pdf	
	g2f_2020_soil_data.csv	.csv	
	z._2020_supplemental_info	directory	
	g2f_2020_supplemental_information.pdf	.pdf	
	g2f_2020_agronomic_information.csv	.csv	
g2f_2020_cooperators_list.csv	.csv		
g2f_2020_field_metadata.csv	.csv		
2021 Planting Season	a._2021_phenotypic_data	directory	CyVerse [4]
	g2f_2021_data_description.pdf	.pdf	(https://doi.org/10.25739/5ae3-sw62)
	g2f_2021_phenotypic_clean_data.csv	.csv	
	g2f_2021_phenotypic_raw_data.csv	.csv	
	b._2021_weather_data	directory	
	g2f_2021_weather_data_description.pdf	.pdf	
	g2f_2021_weather_readMe.txt	.txt	
	g2f_2021_weather_raw.csv	.csv	
	g2f_2021_weather_cleaned.csv	.csv	
	c._2021_soil_data	directory	

g2f_2021_soil_data_description.pdf	.pdf
g2f_2021_soil_data.csv	.csv
z._2021_supplemental_info	directory
g2f_2021_supplemental_information.pdf	.pdf
g2f_2021_agronomic_information.csv	.csv
g2f_2021_cooperators_list.csv	.csv
g2f_2021_field_metadata.csv	.csv

Limitations

These datasets contain missing data. Missing data includes data not reported by collaborators or erroneous data as specified on the readMe and description files. Genotypic data is not included in this release.

Locations that did not collect data due to the complete loss of the experiment are listed on the cooperator list.

Abbreviations

G2F, Genomes to Fields. DH, doubled-haploid. G x E, Genotype by environment.

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and materials

The data described in this Data note can be freely and openly accessed on CyVerse at <https://doi.org/10.25739/hzsz-a865> (2020 Field Season) and <https://doi.org/10.25739/5ae3-sw62> (2021 Field Season).

Competing interests

The authors declare that they have no competing interests.

Funding

We gratefully acknowledge support from: National Corn Growers Association, Iowa Corn Promotion Board, Georgia Corn Commission, Nebraska Corn Board, Ohio Corn Marketing Program, Corn Marketing Program of Michigan, Texas Corn Producers Board, University of Göttingen startup funds, USDA-ARS, and USDA Germplasm Enhancement of Maize program.

Authors' contributions

Data management team: DCL, ACA, RTA, AP, DLS, MC, JH.

Data contributors: DCL, ACA, RTA, NdL, SK, MCR, JLG, JH, TB, MB, EB, JE, SFG, MAG, CNH, JEK, JM, RM, SCM, JS, RSS, MPS, EES, PT, AT, MT, JW, JW, TW, WX.

Communication: NdL, DE, SK.

The data management team aggregated and curated data resources and made them available to other researchers. Data contributors were responsible for advising on data collection methods, collecting the data, reviewing data collection and curation methods, and the resulting datasets. Communicating authors guided the data management process, including data collection, curation, and distribution, and oversaw the writing of the manuscript. All authors reviewed and contributed to the manuscript and agreed with publication.

Acknowledgements

We gratefully acknowledge contributions from many field managers, crop research coordinators, staff, graduate students, student interns and data collectors including: Dustin Eilert, Marina Borsecnik, Renata Barcelos and Ben Fischer (de Leon/Kaeppeler labs, University of Wisconsin - Madison). Naomi Rodman, Hadden Duling, Nathan Moore, Evonne Pinto, Marielle Quinn, & Kai Sanders (Wallace lab). Jason Brewer (USDA-ARS, Raleigh, NC). Trevor Perla, Paige Coffee, Will Deems, and Amy Deariso (ARS-Tifton, GA). Brian Hearn, Gunnar Isaacs, Buddy Willey, Dheeraj Danthuluri, and Raphael Kim (University of Delaware). Tom Siler (Michigan State University). Katherine Guill, Brady Blanton, Mia Ruppel, Grace Sidberry, Daniel Kick (Washburn lab, USDA-ARS). Amanda Gilbert, Peter Hermanson, and Thomas Hoverstad (University of Minnesota).

References

1. Michel KJ, Lima DC, Hundley H, Singan V, Yoshinaga Y, Daum C, et al. Genetic mapping and prediction of flowering time and plant height in a maize Stiff Stalk MAGIC population. *Genetics*. 2022;221.
2. Genomes to Fields resources. 2023. <https://www.genomes2fields.org/resources/>.
3. G2F Consortium. Genomes to Fields 2020 dataset. CyVerse Data Commons. 2020. <https://doi.org/10.25739/hzsz-a865>.
4. G2F Consortium. Genomes to Fields 2021 dataset. CyVerse Data Commons. 2021. <https://doi.org/10.25739/5ae3-sw62>.