2018–2019 field seasons of the Maize Genomes to Fields (G2F) G x E project

Dayane Cristina Lima, Alejandro Castro Aviles, Ryan Timothy Alpers, Bridget A Mcfarland, Shawn Kaeppler, David Ertl, Maria Cinta Romay, Joseph L Gage, James Holland, Timothy Beissinger, et al.

▶ To cite this version:

Dayane Cristina Lima, Alejandro Castro Aviles, Ryan Timothy Alpers, Bridget A Mcfarland, Shawn Kaeppler, et al.. 2018–2019 field seasons of the Maize Genomes to Fields (G2F) G x E project. BMC Genomic Data, 2023, 24 (1), pp.29. 10.1186/s12863-023-01129-2 . hal-04175989

HAL Id: hal-04175989

https://hal.inrae.fr/hal-04175989v1

Submitted on 2 Aug 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution 4.0 International License

Lima *et al. BMC Genomic Data* (2023) 24:29
https://doi.org/10.1186/s12863-023-01129-2

BMC Genomic Data

DATA NOTE **Open Access**

2018–2019 field seasons of the Maize Genomes to Fields (G2F) G x E project

Dayane Cristina Lima[1*], Alejandro Castro Aviles[2], Ryan Timothy Alpers[1], Bridget A. McFarland[3], Shawn Kaeppler[1], David Ertl[4], Maria Cinta Romay[5], Joseph L. Gage[6], James Holland[7], Timothy Beissinger[8], Martin Bohn[9], Edward Buckler[10], Jode Edwards[11], Sherry Flint-Garcia[12], Candice N. Hirsch[13], Elizabeth Hood[14], David C. Hooker[15], Joseph E. Knoll[16], Judith M. Kolkman[17], Sanzhen Liu[18], John McKay[19], Richard Minyo[20], Danilo E. Moreta[17], Seth C. Murray[21], Rebecca Nelson[22], James C. Schnable[23], Rajandeep S. Sekhon[24], Maninder P. Singh[25], Peter Thomison[26], Addie Thompson[25], Mitchell Tuinstra[27], Jason Wallace[28], Jacob D. Washburn[12], Teclemariam Weldekidan[29], Randall J. Wisser[29,30], Wenwei Xu[31] and Natalia de Leon[1]

Abstract

Objectives This report provides information about the public release of the 2018–2019 Maize G X E project of the Genomes to Fields (G2F) Initiative datasets. G2F is an umbrella initiative that evaluates maize hybrids and inbred lines across multiple environments and makes available phenotypic, genotypic, environmental, and metadata information. The initiative understands the necessity to characterize and deploy public sources of genetic diversity to face the challenges for more sustainable agriculture in the context of variable environmental conditions.

Data description Datasets include phenotypic, climatic, and soil measurements, metadata information, and inbred genotypic information for each combination of location and year. Collaborators in the G2F initiative collected data for each location and year; members of the group responsible for coordination and data processing combined all the collected information and removed obvious erroneous data. The collaborators received the data before the DOI release to verify and declare that the data generated in their own locations was accurate. ReadMe and description files are available for each dataset. Previous years of evaluation are already publicly available, with common hybrids present to connect across all locations and years evaluated since this project's inception.

Keywords Maize, Genotype by environment, Phenotype, Variable environments, Grain yield

Objective

Maize (*Zea mays* subsp. *mays* L.) plays an important role in the global economy. As a crop, it displays a variety of uses such as food, feed, and fuel. At the same time, and due to its versatility and relevance, maize has been widely studied. The Genomes to Fields (G2F) is a collaborative initiative involving scientists from the public sector that support growers, consumers, and society. G2F researchers generate phenotypic, genotypic, environmental, and metadata datasets to facilitate the understanding of the potential and challenges of maize production in different environments.

Individual genotype performances differ across environments, and the magnitude of this difference dictates the importance of the Genotype by Environment (G×E) interaction. Understanding and harnessing G×E

---

*Correspondence:
Dayane Cristina Lima
dclima@wisc.edu
Full list of author information is available at the end of the article

© The Author(s) 2023. **Open Access** 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 licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence 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 licence, 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.

interactions improves the efficiency in the use and allocation of resources, and it facilitates the identification of genotypes with higher stability across a range of locations, the identification of locations where the effect of G x E is minimized, and the identification of mechanisms affecting the differential response of phenotypes to variable environments. Furthermore, advances in our understanding of the fundamental components contributing

Table 1 Overview of data files and dataset for 2018 and 2019 planting seasons

Label	Name of data file/data set	File types (Extension)	Data repository and identifier (DOI or accession number)
Data set 1	Evaluation of genetic diversity across the inbreds used by G2F project (WGS skim sequencing)	fastq files (.fq.gz)	NCBI *BioProject* (https://identifiers.org/ncbi/bioproject:PRJNA530187) [4]
Data file 1	README.txt	Text file (.txt)	CyVerse (https://doi.org/10.25739/anqq-sg86) [6]
Data file 2	README.txt	Text file (.txt)	CyVerse (https://doi.org/10.25739/anqq-sg86) [6]
Data file 3	_g2f_2018_hybrid_data_description.pdf	Portable Document Format (.pdf)	CyVerse (https://doi.org/10.25739/anqq-sg86) [6]
Data file 4	g2f_2018_hybrid_data_clean.csv	comma-separated values file (.csv)	CyVerse (https://doi.org/10.25739/anqq-sg86) [6]
Data file 5	g2f_2018_hybrid_data_raw.csv	comma-separated values file (.csv)	CyVerse (https://doi.org/10.25739/anqq-sg86) [6]
Data file 6	raw_cleaning_readme.txt	Text file (.txt)	CyVerse (https://doi.org/10.25739/anqq-sg86) [6]
Data file 7	README_weather.txt	Text file (.txt)	CyVerse (https://doi.org/10.25739/anqq-sg86) [6]
Data file 8	_g2f_2018_weather_data_description.pdf	Portable Document Format (.pdf)	CyVerse (https://doi.org/10.25739/anqq-sg86) [6]
Data file 9	g2f_2018_weather_clean.csv	comma-separated values file (.csv)	CyVerse (https://doi.org/10.25739/anqq-sg86) [6]
Data file 10	g2f_2018_weather_raw.csv	comma-separated values file (.csv)	CyVerse (https://doi.org/10.25739/anqq-sg86) [6]
Data file 11	weather_cleaning_readme.txt	Text file (.txt)	CyVerse (https://doi.org/10.25739/anqq-sg86) [6]
Data file 12	Indigo_2018_Soil_Data.pdf	Portable Document Format (.pdf)	CyVerse (https://doi.org/10.25739/anqq-sg86) [6]
Data file 13	_g2f_2018_soil_description.pdf	Portable Document Format (.pdf)	CyVerse (https://doi.org/10.25739/anqq-sg86) [6]
Data file 14	g2f_2018_soil_data.csv	comma-separated values file (.csv)	CyVerse (https://doi.org/10.25739/anqq-sg86) [6]
Data file 15	G2F_PHG_minreads1_Mo44_PHW65_MoG_assemblies_14112019_filtered_plusParents.vcf	variant call format file (.vcf)	CyVerse (https://doi.org/10.25739/anqq-sg86) [6]
Data file 16	G2F_PHG_minreads1_Mo44_PHW65_MoG_assemblies_14112019_filtered_plusParents_description.pdf	Portable Document Format (.pdf)	CyVerse (https://doi.org/10.25739/anqq-sg86) [6]
Data file 17	G2F_PHG_minreads1_Mo44_PHW65_MoG_assemblies_14112019_filtered_plusParents_sampleDecoder.txt	Text file (.txt)	CyVerse (https://doi.org/10.25739/anqq-sg86) [6]
Data file 18	README_G2F_2020-03–13.txt	Text file (.txt)	CyVerse (https://doi.org/10.25739/anqq-sg86) [6]
Data file 19	README_Genotypic.txt	Text file (.txt)	CyVerse (https://doi.org/10.25739/anqq-sg86) [6]
Data file 20	g2f_2018_agronomic information.csv	comma-separated values file (.csv)	CyVerse (https://doi.org/10.25739/anqq-sg86) [6]
Data file 21	g2f_2018_cooperators_list.csv	comma-separated values file (.csv)	CyVerse (https://doi.org/10.25739/anqq-sg86) [6]
Data file 22	g2f_2018_field_metadata.csv	comma-separated values file (.csv)	CyVerse (https://doi.org/10.25739/anqq-sg86) [6]
Data file 23	g2f_2018_supplemental_information.txt	Text file (.txt)	CyVerse (https://doi.org/10.25739/anqq-sg86) [6]
Data file 24	g2f_planting_season_2019_readMe.txt	Text file (.txt)	CyVerse (https://doi.org/10.25739/t651-yy97) [7]
Data file 25	g2f_2019_phenotypic_clean_data.csv	comma-separated values file (.csv)	CyVerse (https://doi.org/10.25739/t651-yy97) [7]
Data file 26	g2f_2019_phenotypic_data_description.pdf	Portable Document Format (.pdf)	CyVerse (https://doi.org/10.25739/t651-yy97) [7]
Data file 27	g2f_2019_phenotypic_data_read_me.txt	Text file (.txt)	CyVerse (https://doi.org/10.25739/t651-yy97) [7]
Data file 28	g2f_2019_phenotypic_raw_data.csv	comma-separated values file (.csv)	CyVerse (https://doi.org/10.25739/t651-yy97) [7]
Data file 29	2019_weather_cleaned.csv	comma-separated values file (.csv)	CyVerse (https://doi.org/10.25739/t651-yy97) [7]
Data file 30	2019_weather_raw.csv	comma-separated values file (.csv)	CyVerse (https://doi.org/10.25739/t651-yy97) [7]
Data file 31	g2f_2019_weather_data_description.pdf	Portable Document Format (.pdf)	CyVerse (https://doi.org/10.25739/t651-yy97) [7]
Data file 32	g2f_2019_weather_readMe.txt	Text file (.txt)	CyVerse (https://doi.org/10.25739/t651-yy97) [7]
Data file 33	g2f_2019_soil_data.csv	comma-separated values file (.csv)	CyVerse (https://doi.org/10.25739/t651-yy97) [7]
Data file 34	g2f_2019_soil_data_description.pdf	Portable Document Format (.pdf)	CyVerse (https://doi.org/10.25739/t651-yy97) [7]
Data file 35	g2f_2019_agronomic_information.csv	comma-separated values file (.csv)	CyVerse (https://doi.org/10.25739/t651-yy97) [7]
Data file 36	g2f_2019_cooperators_list.csv	comma-separated values file (.csv)	CyVerse (https://doi.org/10.25739/t651-yy97) [7]
Data file 37	g2f_2019_field_metadata.csv	comma-separated values file (.csv)	CyVerse (https://doi.org/10.25739/t651-yy97) [7]
Data file 38	g2f_2019_supplemental_information.pdf	Portable Document Format (.pdf)	CyVerse (https://doi.org/10.25739/t651-yy97) [7]

to the differential response of plants to environmental cues will also improve genomic and phenotypic predictabilities for traits of interest. Therefore, this data release provides a unique resource of combined agronomic, phenological, and morphological information to dissect G × E interaction.

In the 2018 and 2019 experiments, 1153 publicly available hybrids were evaluated through a network of collaborators in 32 different locations. The main group of hybrids was produced by the cross of doubled-haploid (DH) inbred lines from a collection of three biparental populations that share one parent in common (PHW65) and PHN11, Mo44, and MoG as the alternative parent, to two ex-PVP inbred testers, LH195 in Midwest to Southern locations, and PHT69 in Northern locations.

Data description

The 2018 and 2019 datasets are publicly available via CyVerse/iPlant and structured as described in Table 1. Briefly, the datasets included here are:

- **Phenotypic dataset:** Phenotypic measurements that follow a standard set of instructions, available in the G2F webpage [1]. 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 estimated grain yield. Raw data and quality-controlled data are reported. Out of range observations were set to missing following the rules described in the readMe and data description files.
- **Genotypic dataset:** Inbred parents of the tested hybrids were genotyped using the Practical Haplotype Graph (PHG) [2, 3]. The data is minimally filtered, allowing the public to perform their own quality control steps prior to using it. The raw sequencing reads are available under BioProject ID PRJNA530187 [4]. The code used to create the genotypic data is also available at https://bitbucket.org/bucklerlab/g2f_2018_phg_genotyping/src/master/.
- **Environmental dataset:** WatchDog 2700 weather stations (Spectrum Technologies) were placed at each field site. Data was collected at 30-min intervals 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 in the G2F webpage [5].
- **Soil dataset:** Each field location collected soil samples that represent the experiment field. Collaborators follow instructions available on the G2F webpage for sample collection [5].
- **Supplemental dataset:** Supplemental information consists of metadata (any field-level data collected at planting, in season, and/or at harvest), agronomic information (list of pesticides, nutrients, and irrigation applied), and cooperator list (collaborators responsible for the field locations in 2018 and 2019).

Limitations

These datasets contain missing data. Missing data includes data not reported by collaborators or erroneous data as determined by data description files. In 2019, some locations had pedigree information set to missing due to packaging problems and only plot number was reported in the phenotypic dataset to reduce misinterpretation.

Abbreviations

G2F	Genomes to Fields
DOI	Digital Object Identifier
DH	Doubled-haploid
G x E	Genotype by environment

Acknowledgements

We gratefully acknowledge contributions from many field managers and data collectors including: Dustin Eilert, Marina Borsecnik, Rachel Perry, Renata Barcelos and Ben Fischer (de Leon/Kaeppler labs, University of Wisconsin—Madison). Amanda Gilbert (Hirsch lab, University of Minnesota). Christine Smith, Brandi Sigmon, Connor Pedersen, Nathaniel Pester, Isaac Stevens (Schnable Lab, University of Nebraska-Lincoln). Trevor Perla, Amy Deariso, Paige Coffee, Steven Hughes, and C.J. Dudley (USDA Tifton). Naomi Rodman, Spencer Caro, Coleman Grindle, Allison McCabe, Samuel Morris, and Bamidele Sangoyomi (Wallace lab—University of Georgia). William Widdicombe and Linsey Newton (Singh and Thompson Labs, Michigan State University). Susan Melia-Hancock and Jim Elder (Flint-Garcia Lab, USDA-ARS, Columbia MO). Emmalea Ernest and Victor Green (University of Delaware). Colby Bass and Regan Lindsey (Texas A&M University System). Kyle Evans, Kirsten Hein, Anne Howard, Jack Mullen, Patrick Woods (McKay Lab, Colorado State University). Dietrich Kaufmann (Beissinger Lab). Christina Poudyal, Kevin Silverstein, Anna Rogers, Luis Samayoa, Tyson Swetnam. In addition, we gratefully acknowledge contributions from numerous staff indirectly involved in the project, graduate students, and student workers at many locations.

Authors' contributions

Data management team: DCL, ACA, RTA, BAM, MCR, JLG, JH, JE, DE, JDW. Data contributors: DCL, ACA, RTA, NdL, SK, MCR, JLG, JH, TB, MB, EB, SFG, JE, CNH, EH, DCH, JEK, JMK, SL, JM, RM, DEM, SCM, RN, JCS, RSS, MPS, PT, AT, MT, JW, JDW, TW, RJW, WX. Communication: NdL, DE, SK. The data management team aggregated, curated, and made available data resources. Data contributors advised on data collection methods, collected the data, and reviewed data collection and curation methods as well as datasets. Communicating authors guided data collection, curation, and distribution. All authors reviewed the manuscript. All authors read and approved the final manuscript.

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.

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/anqq-sg86 (2018 Field Season) and https://doi.org/10.25739/t651-yy97 (2019 Field Season). Please see Table 1 and references list for details and links to the data.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

[1]Department of Agronomy, University of WI – Madison, Madison, WI 53706, USA. [2]BASF Vegetable Seeds, Brooks, OR 97305, USA. [3]Panama-USA Commission for the Eradication and Prevention of Screwworm (COPEG), USDA-APHIS-IS, Pacora, Panama. [4]Iowa Corn Promotion Board, Johnston, IA 50131, USA. [5]Institute for Genomic Diversity, Cornell University, Ithaca, NY 14853, USA. [6]Department of Crop and Soil Sciences, North Carolina State University, Raleigh, NC 27695, USA. [7]USDA-ARS Plant Science Research Unit, Raleigh, NC 27606, USA. [8]Department of Crop Science, University of Göttingen Center for Integrated Breeding Research, Carl-Sprengel-Weg 1, 37075 Göttingen, Germany. [9]University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA. [10]USDA-ARS and Cornell University, Ithaca, NY 14853, USA. [11]USDA ARS CICGRU, 716 Farmhouse Ln, Ames, IA 50011-1051, USA. [12]USDA-ARS, Columbia, MO 65211, USA. [13]Department of Agronomy and Plant Genetics, University of Minnesota, St Paul, MN 55108, USA. [14]College of Agriculture, Arkansas Biosciences Institute, Arkansas State University, Jonesboro, AR 72404, USA. [15]Department of Plant Agriculture, University of Guelph, Ridgetown Campus, Ridgetown, ON, Canada. [16]USDA-ARS Crop Genetics and Breeding Research Unit, Tifton, GA 31793, USA. [17]School of Integrative Plant Science, Cornell University, Ithaca, NY 14850, USA. [18]Department of Plant Pathology, Kansas State University, Manhattan, KS 66503, USA. [19]Department of Agricultural Biology, Colorado State University, Fort Collins, CO 80523, USA. [20]Department of Horticulture and Crop Science, Ohio State University College of Food, Agricultural, and Environmental Sciences, Wooster, OH 44691, USA. [21]Department of Soil and Crop Sciences, Texas A&M University, College Station, TX 77843, USA. [22]Cornell University, Ithaca, NY 14853, USA. [23]Department of Agronomy and Horticulture, University of Nebraska-Lincoln, Lincoln, NE 68588, USA. [24]Department of Genetics and Biochemistry, Clemson University, Clemson, SC 29634, USA. [25]Department of Plant, Soil and Microbial Sciences, Michigan State University, East Lansing, MI 48824, USA. [26]Ohio State University, Columbus, OH 43210, USA. [27]Department of Agronomy, Purdue University, West Lafayette, IN 49707, USA. [28]Department of Crop & Soil Sciences, University of Georgia, Athens, GA 30602, USA. [29]Department of Plant and Soil Sciences, University of Delaware, Newark, DE 19716, USA. [30]Laboratoire d'Ecophysiologie Des Plantes Sous Stress Environmentaux, INRAE, 34060 Montpellier, France. [31]Texas A&M University, College Station, TX 77843, USA.

Received: 12 October 2022 Accepted: 16 May 2023
Published online: 25 May 2023

References

1. Genomes to Fields. 2022. https://www.genomes2fields.org. Accessed 10 Oct 2022.
2. Bradbury PJ, Casstevens T, Jensen SE, Johnson LC, Miller ZR, Monier B, et al. The Practical Haplotype Graph, a platform for storing and using pangenomes for imputation. Bioinformatics. 2022;38(15):3698–702. https://doi.org/10.1093/bioinformatics/btac410.
3. Franco JAV, Gage JL, Bradbury PJ, Johnson LC, Miller ZR, Buckler ES, et al. A Maize Practical Haplotype Graph Leverages Diverse NAM Assemblies. bioRxiv. 2020. https://doi.org/10.1101/2020.08.31.268425.
4. Evaluation of genetic diversity across the inbreds used by G2F project (WGS skim sequencing). BioProject. 2022. https://identifiers.org/ncbi/bioproject:PRJNA530187.
5. Genomes to Fields resources. 2022. https://www.genomes2fields.org/resources/. Accessed 10 Oct 2022.
6. G2F Consortium. Genomes to Fields 2018 Data Set. CyVerse Data Commons. 2018. https://doi.org/10.25739/anqq-sg86.
7. G2F Consortium. Genomes to Fields 2019 dataset. CyVerse Data Commons. 2019. https://doi.org/10.25739/t651-yy97.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

BMC