



December 2018: Report for the Corn Germplasm Committee
Prepared by Jack Gardiner, Margaret Woodhouse, Carson Andorf, and the MaizeGDB team

Maize Genomes Update

In the last year, a few new maize genomes have been sequenced and released: the Mo17 genome sequenced by Jinshing Lai's group at the Chinese Agricultural University (CAU) in Beijing, and two European Flint lines, DK105 and PE0075, by Eva Bauer's group at the Technical University of Munich (TUM) (<https://maizegdb.org/genome/assemblies/overview>). MaizeGDB is currently hosting the CAU Mo17, and is in the process of preparing the new European Flints for our December update. The 25 NAM (Nested Association Mapping) founder line genomes are currently being sequenced and assembled by the NAM Sequencing Consortium (Matt Hufford of ISU, Kelly Dawe of UGA, and Doreen Ware of CSHL), and will be released in the upcoming months, along with gene model annotations and RNA-seq data. MaizeGDB is working on a plan to host these high-quality genomes in the most efficient way, including connecting gene model, functional, phenotype, and germplasm data among the founder lines using a pan-genomic framework. MaizeGDB continues to support the reference assembly B73 v4.

Maize Genome Stewardship Update

While continuing to require that researchers submit their genomes to GenBank before submitting their genomes to MaizeGDB, we are also developing a plan to determine the level of support MaizeGDB will provide each genome. Support levels include (1) Full: genomes will have browser tracks, blast targets, and gene model pages for all gene models, as well as functional or phenotype data where applicable; (2) Partial: genomes will have browser tracks and/or blast targets; (3) Limited: a link to genome downloads will be provided. Standards to determine level of support will include: (1) to what extent the genome adheres to FAIR (Findable, Accessible, Interoperable, and Reusable) practices (<http://info.hsls.pitt.edu/updatereport/?p=9493>); (2) how unique the genome is diversity-wise compared to genomes currently being hosted in MaizeGDB; (3) genome quality. This plan is still in the developmental stages at the current time.

Data Highlights for 2018

- MaizeGDB has added 124 RNA-seq samples from two different experiments (Stelpflug 2015 and Walley 2016) to the maize B73 v4 genome browser so that users can visualize RNA expression within a selected chromosomal interval.
- MaizeGDB has added the tool qTeller to view, query, and compare gene expression data in B73 v4.
- MaizeGDB has developed the tool SNPiversity to view and query large-scale maize diversity data.
- MaizeGDB has developed the tool Pedigree Viewer to visualize pedigree data in graphical formats.
- Blast targets, downloads, metadata, and browser tracks for the Mo17 genome by the Chinese Agricultural University in Beijing, China is now available at MaizeGDB.

- Blast targets, downloads, and metadata for the new European Flint lines DK105 and PE0075 from TUM will be available in the December update.
- MaizeGDB is working closely with the NAM Sequencing Consortium to submit genomes to GenBank and to decide the best way to host the NAM genomes in the database.

Tool Highlights

In the last year, MaizeGDB has developed, implemented, and published multiple new tools for improved access and visualization of data. These tools provide the ability to view and query large-scale maize diversity data (SNPiversity), visualize pedigree data in graphical formats (Pedigree Viewer), and view, query, and compare gene expression data (qTeller).

SNPiversity

SNPiversity allows maize researchers to select a customized set of maize lines and a genomic region, and visualize DNA variations for that genomic region. The tool is loaded with datasets from Panzea that contain approximately 1 million regions of DNA variation for over 17,000 public maize lines. Genotype data for thousands of maize lines and individuals that represent the broad diversity represented by the *Zea* genus (i.e., maize and its near relatives) are also available as downloads through our Diversity Data center at <https://www.maizegdb.org/diversity>.

Pedigree Viewer

The Pedigree Viewer is a web-based pedigree viewer for maize. It allows users to build a network around a line, find the shortest path between two lines, build a network around the least common ancestor of two stocks, and filter pedigree networks by state, developer, source, and country. The Pedigree Viewer at MaizeGDB has a dynamically-generated pedigree network of 4,706 maize lines and 5,487 relationships. The tool has also been integrated into MaizeGDB's stock pages to show interactive images of pedigree trees. The Pedigree Viewer is customizable and allows users to upload their own data. This tool will help maize breeders identify appropriate maize lines to use in breeding projects which will improve available germplasm for researchers and farmers. https://maizegdb.org/breeders_toolbox

qTeller

qTeller (<http://www.qteller.com>) is an RNA-seq processing pipeline and modular web interface that has been used to study quantitative expression variation in various species. qTeller draws RNA expression graphs for multiple RNA-seq datasets in real time for any gene of interest, and can be used to compare two genes at once by drawing a dot plot of relative expression for each RNA-seq dataset. MaizeGDB now hosts a version of qTeller (<https://qteller.maizegdb.org>) that currently has RNA-seq datasets from six different publications for B73, covering over 150 different tissues, conditions, and/or time points, and will soon host RNA-seq datasets for the inbred lines W22, Mo17, and others.

MaizeMine

MaizeMine is a data mining warehouse at MaizeGDB (<http://maizemine.maizegdb.org>) that accelerates genomic analysis by enabling researchers without scripting skills to create and export customized annotation datasets in a variety of formats that can be merged with their own research data for use in downstream analyses. MaizeMine uses the InterMine data warehousing system to fully integrate genomic sequences from both the B73_RefGen_v3 and _v4 genome assemblies. MaizeMine provides a variety of simple and sophisticated search tools to MaizeGDB stakeholders. In addition to data sets added previously, we have added 10 additional data sets on transcription start sites, insertional and EMS mutations, and the Illumina 50K SNP

markers. Future work with MaizeMine will include support for additional B73 assemblies as they become available with cross indexing to older B73 assemblies.

MaizeDIG

MaizeDIG (<http://maizedig.maizegdb.org>) is based on the BioDIG database for linking images to genomic data. MaizeDIG is configured to access 2700 images contained within the MaizeGDB database as well as their corresponding gene models. A major feature of MaizeDIG is the creation of a novel data track that enables a pop-up of phenotypic images corresponding to their gene model to be shown within the context of the MaizeGDB genome browser. All genome assemblies hosted at MaizeGDB are accessed by the MaizeDIG database. Future work on MaizeDIG will include working with the Maize Genetics Community to include new images which have been associated to a gene model.

MaizeGDB Outreach Activities

MaizeGDB organized a series of four workshops at the 60th Annual Maize Genetics Conference in St. Malo, France (<https://community.maizegdb.org/workshops/workshops.php>). MaizeGDB, in collaboration with Gramene (www.gramene.org), also organized the Agricultural Database booth at the 2018 Plant and Animal Genome (PAG) Meeting in San Diego where personnel from the various agricultural databases can present informational materials to PAG attendees and give hands on, in person tutorials.

The AgBioData Consortium is a group of scientists from 30 different plant and animal genetic, genomic and breeding databases. MaizeGDB is a founding member of AgBioData and Ethalinda Cannon and Lisa Harper are members of the Steering Committee. The mission is to work together to ensure standards and best practices for acquisition, display and retrieval of genomic, genetic and breeding (GGB) data, and to promote the FAIR data principles (<https://www.force11.org/group/fairgroup/fairprinciples>). In 2018, Lisa Harper was lead author on an extensive open access article on standards and best practices for GGB databases (<https://academic.oup.com/database/article/2018/1/bay088/5096675?searchresult=1>). To fund workshops, outreach, and other endeavors, the AgBioData Steering Committee has recently applied for a USDA NIFA Food and Agriculture Cyberinformatics and Tools Initiative grant titled "FACT: AgBioData - A Coordinated, Collaborative and Innovative Network of Genomic, Genetic and Breeding Databases for Enhanced Agricultural Research Outcomes".

Maize Working Group Meeting

The Maize Working Group met on September 26th, 2018 with Dorrie Main acting as the new WG Chair (see acknowledgements for other members). A detailed report on MaizeGDB activities was provided to the WG prior to the meeting and this and past reports can be found here: https://maizegdb.org/about/working_group In addition, several charge questions were provided to assist them in providing specific guidance on challenging issues that MaizeGDB has identified.

Community Leadership Roles

- USDA-ARS Executive Scientific Advisory Committee – C. Andorf
- USDA-ARS CERES High Performance Computing Policy Committee – J. Portwood
- USDA-ARS Database Committee - J. Portwood
- Maize Genetics Executive Committee - C. Andorf (ex-officio)
- Maize Genetics Conference Steering Committee – C. Andorf (ex-officio)
- AgBioDatabase Consortium Steering Committee – L. Harper (chair) and E. Cannon
- Maize Nomenclature Committee – L. Harper and E. Cannon

Staffing at MaizeGDB

Carson Andorf (USDA-ARS, 1.0 FTE) is the lead scientist for MaizeGDB. Maggie Woodhouse (Iowa State University, 1.0 FTE) is a scientist leading the genomics efforts at MaizeGDB. Lisa Harper (USDA-ARS, 0.5 FTE) and Jack Gardiner (University of Missouri, 1.0 FTE) are curators. Retired curator Mary Schaeffer continues to work part time (0.2 FTE) at MaizeGDB. Ethy Cannon (USDA-ARS, 0.5 FTE) is a bioinformatics engineer. John Portwood (USDA-ARS, 1.0 FTE) is the database administrator and lead developer on the web interface. Current students funded through MaizeGDB are Nancy Manchanda (B014 assembly and annotation), Kyoung Tak Cho (MaizeDIG and machine learning applications), Sagnik Banerjee (Gene prediction assessment) and Mandy Dietze (curation). Two employees have moved on from MaizeGDB in the past year. David Schott has graduated with a degree in computer science and works for Microsoft. Jesse Walsh completed his postdoctoral fellowship and works as a research associate at the Mayo Clinic. The two vacant permanent positions at MaizeGDB have been advertised and are expected to be filled in the next couple months. These positions are for a Computational Biologist (Scientist) and Web Database Programmer (IT-Specialist).

Publications for 2018

Portwood JL II, Woodhouse MR, Cannon EK, Gardiner JM, Harper LC, Schaeffer ML, Walsh JR, Sen TZ, Cho KT, Schott DA, Braun BL, Dietze M, Dunfee B, Elisk CG, Manchanda N, Coe E, Sachs M, Stinard P, Tolbert J, Zimmerman S, Andorf CM. (2018) MaizeGDB 2018: the maize multi-genome genetics and genomics database. *Nucl Acids Research* 08 November 2018 doi: 10.1093/nar/gky1046.

Alkhalifah N, Campbell DA, Falcon CM, Gardiner JM, Miller ND, Romay MC, Walls R, Walton R, Yeh CT, Bohn M, Bubert J, Buckler ES, Ciampitti I, Flint-Garcia S, Gore MA, Graham C, Hirsch C, Holland JB, Hooker D, Kaeppeler S, Knoll J, Lauter N, Lee EC, Lorenz A, Lynch JP, Moose SP, Murray SC, Nelson R, Rocheford T, Rodriguez O, Schnable JC, Scully B, Smith M, Springer N, Thomison P, Tuinstra M, Wisser RJ, Xu W, Ertl D, Schnable PS, De Leon N, Spalding EP, Edwards J, Lawrence-Dill CJ. (2018) Maize Genomes to Fields: 2014 and 2015 field season genotype, phenotype, environment, and inbred ear image datasets. *BMC Res Notes*. 2018 Jul 9;11(1):452. doi: 10.1186/s13104-018-3508-1.

Harper L, Campbell J, Cannon EK, Jung S, Main D, Poelchau M, Walls RL, Andorf CM, Arnaud E, Berardini, Birkett C, Cannon S, Carson J, Cooper L, Dunn N, Elisk C, Farmer A, Ficklin S, Grant D, Grau E, Hendon N, Hu Z, Humann J, Jaiswal P, Jonquet C, Laporte MA, Larmande P, Lazo G, McCarthy F, Menda N, Mungall C, Munoz-Torres M, Naithani S, Nelson R, Neddill D, Park C, Reecy J, Reiser L, Sanderson LA, Sen TZ, Staton M, Subramaniam S, Karey Tello-Ruiz M, Unda V, Unni D, Wang L, Ware D, Wegrzyn J, Williams J, Woodhouse M. (2018) AgBioData Consortium Recommendations for Sustainable Genomics and Genetics Databases for Agriculture. Database. Volume 2018.

Schott DA, Vinnakota AG, Portwood JL, Andorf CM, Sen TZ. (2018) SNPiversity: a web-based tool for visualizing diversity. Database. Volume 2018.

Siegel ZD, Zhou N, Zarecor S, Lee N, Campbell DA, Andorf CM, Nettleton D, Lawrence-Dill CJ, Ganapathysubramanian B, Friedberg I, Kelly JW. (2018) Crowdsourcing Image Analysis for Plant Phenomics to Generate Ground Truth Data for Machine Learning. *PLOS Computational Biology*.

Springer NM, Anderson SN, Andorf CM, Ahern KR, Bai F, Barad O, Barbazuk WB, Bass HW, Baruch K, Ben-Zvi G, Buckler E, Bukowski R, Campbell MS, Cannon EKS, Chomet P, Dawe RK, Davenport R, Dooner HK, Du LH, Du C, Easterling KA, Gault CM, Guan JC, Hunter CT III, Jander G, Jiao YP, Koch KE, Kol G, Kollner TG, Kudo T, Li Q, Lu F, Mayfield-Jones D, Mei WB, McCarty DR, Noshay JM, Portwood JL, Ronen G, Settles AM, Shem-Tov D, Shi JH, Soifer I, Stein JC, Stitzer MC, Suzuki M, Vera DL, Vollbrecht E, Vrebalov JT, Ware DH, Wei S, Wimalanathan K, Woodhouse MHR, Xiong WW, Brutnell TP. (2018). The maize W22 genome provides a foundation for functional genomics and transposon biology. *Nature Genetics*. doi: 10.1038/s41588-018-0158-0

Wimalanathan K, Friedberg I, Andorf CM, Lawrence-Dill CJ. (2018) Maize GO Annotation- Methods, Evaluation, and Review (maize-GAMER). *Plant Direct*. Vol 2, Issue 4. e00052.

Presentations for 2018

Harper LC, Cho KT, Portwood J, Gardiner JM, Schaeffer ML, Woodhouse MR, Cannon EK, Lawrence CJ, Friedberg I, Andorf CM (speaker). How phenotype curation has co-evolved with genomic representations. 5th International Plant Phenotyping Symposium, Adelaide, Australia in 2018.

Woodhouse MR, Portwood JL II, Cannon EK, Gardiner JM, Harper LC, Schaeffer ML, Walsh JR, Andorf CM (speaker). MaizeGDB: 2018 Update. National Corn Growers Corn Congress 2018, Washington, D.C. in 2018.

Woodhouse MR, Portwood JL II, Cannon EK, Gardiner JM, Harper LC, Schaeffer ML, Walsh JR, Andorf CM (speaker). MaizeGDB: 2018 Update. National Corn Growers State and National Staff Meeting, St. Louis, MO in 2018.

Cho KT (speaker), Portwood J, Harper LC, Gardiner JM, Lawrence CJ, Friedberg I, Andorf CM. MaizeDIG: A mechanism for connecting gene models to phenotypes at MaizeGDB. Short Talk at 26th Annual Conference Intelligent Systems for Molecular Biology (ISMB), Chicago, IL in 2018.

Wimalanathan K (speaker), Friedberg I, Andorf CM, Lawrence-Dill C. Maize GO Annotation - Methods, Evaluation, and Review (maize-GAMER). Short Talk at 6th Annual Conference Intelligent Systems for Molecular Biology (ISMB), Chicago, IL in 2018.

Wimalanathan K (speaker), Friedberg I, Andorf CM, Lawrence-Dill C. Maize - GO Annotation Methods Evaluation and Review (Maize-GAMER). Short Talk at 60th Annual Maize Genetics Conference (MGC 2018), St. Malo, France.

Woodhouse M (speaker), Cannon EK, Andorf CM. The Importance of Getting Genome Assemblies into Genbank, and How to Do It. Short Talk during the Big Data: Manage Your Data Before Your Data Kills You Workshop at Plant and Animal Genome Conference, San Diego, CA in 2018.

Elsik, C (Speaker) MaizeMine Workshop within the Maize Tools and Resources Session at the 60th Annual Maize Genetics Conference (MGC2018), St Malo, France.

Woodhouse M (speaker). How the Maize Genome Database Helps Guide Data Management Best Practices in the Maize Community. Short Talk during the Challenges and Opportunities in

Plant Science Data Management - an International Workshop at Plant and Animal Genome Conference, San Diego, CA in 2018.

Elsik, C (speaker). MaizeMine: A Data Mining Warehouse for MaizeGDB. Short Talk during the Digital Tools and Resources Session at Plant and Animal Genome Conference, San Diego, CA in 2018.

Acknowledgements

MaizeGDB is funded by the USDA-ARS. MaizeGDB would like to thank the National Program staff, particularly, Jack Okamoto, Peter Bretting, and Jose Costa. MaizeGDB would like to acknowledge the guidance provided by the MaizeGDB Working group: D Main (Chair), A Barkan, Q Dong, D Jackson, T Lubberstedt, C Hirsch, M Sachs (*ex officio*), M Settles, and J Schnable, the Maize Genetics Executive Committee: J Yu (Chair), E Buckler, N de Leon, D Jackson, S Kaepler, K Koch, K Newton, P Schnable, and M Warburton; the Maize Nomenclature Committee: M Sachs (Chair), E Buckler, H Dooner (retired), C Du, T Kellogg, and P Stinard.