December 2017: Report for the Corn Germplasm Committee
Prepared by Jack Gardiner, Carson Andorf, and the MaizeGDB team

GENOME ASSEMBLY STEWARDSHIP

One of the main priorities at MaizeGDB is to provide genome assembly and annotation stewardship for the maize research community. With falling sequencing costs and improved genome assembly methods, it has become feasible to generate dozens of reference-quality genome assemblies for maize accessions of importance to maize breeders and researchers. MaizeGDB currently hosts information for 10 high-quality genome assemblies (B104, B73, CML247, EP1, F7, Ki11, Mo17, NC350, PH207, and W22) (https://maizegdb.org/genome/assemblies_overview) and has integrated them with data held by MaizeGDB. This enables both exploring individual genomes, and comparing them in sets. In anticipation of more genomes expected in the near future, MaizeGDB developed a set of minimum standards for hosting a new genome assembly, designed templates for collecting essential metadata related to the genome and assembly, enforced naming conventions set out by the maize nomenclature committee, created documentation to help submit genome assemblies to GenBank, and developed a pipeline for loading new assemblies. All of this enables comparative analysis. In addition to bringing in new genome assemblies and providing the research community with a means of improvement, MaizeGDB will continue stewardship of the B73 genome assembly and annotation (https://maizegdb.org/assembly), which is expected to remain the representative reference maize genome assembly for the foreseeable future. Multiple, high-quality genome assemblies and annotations integrated with trait, phenotype, and germplasm data, will improve researchers' ability to conduct trait and germplasm analyses and to identify appropriate germplasm for breeding programs.

Overview of the Genome Assemblies for Maize

Assembly Name	Identifier	Line	Quality	Toronto Agreement	Status
Zm-B104-DRAFT-ISU_USDA-0.1	Zm00007a	B104	Draft genome	yes	Beta version released on October 4th 2016 under Toronto Agreement. Draft pseudomolecule release in December 2016. Full reference genome release Fall 2017.
Zm-B73-REFERENCE-GRAMENE-4.0	Zm00001d	B73	Full reference genome	no	Assembly completed and annotation is nearly complete
Zm-CML247-DRAFT-PANZEA-1.0	Zm00006a	CML247	Draft genome		Assembly and annotation completed
Zm-EP1-REFERENCE-TUM-1.0	Zm00010a	EP1	Reference genome	yes	Assembly completed and annotation in progress
Zm-F7-REFERENCE-TUM-1.0	Zm00011a	F7	Reference genome	yes	Assembly completed and annotation in progress
Zm-Ki11-REFERENCE-GRAMENE-1.0	Zm00012a	Ki11	Reference genome		In progress
Zm-Mo17-REFERENCE-NRGENE-1.0	Zm00005a	Mo17	Reference genome		Completed
Zm-Mo17---1.0	Zm00009a	Mo17	unknown		Near completion
Zm-NC350-REFERENCE-GRAMENE-1.0	Zm00013a	NC350	Reference genome		In progress
Zm-PH207-REFERENCE_NS-UIUC_UMN-1.0	Zm00008a	PH207	Reference genome		Assembly and annotation completed
Zm-W22-REFERENCE-NRGENE-2.0	Zm00004b	W22	Reference genome	yes	Assembly and annotation completed

Click here to learn about maize genome and gene model nomenclature rules.

Genome Assemblies Overview. Overview of the genome assemblies currently publicly listed at MaizeGDB. MaizeGDB hosts maize genomes with metadata, structural and functional annotations, genome browsers, and BLAST targets.

Creation of MaizeMine, a Data Warehouse for Maize

In collaboration with Dr. Chris Elsik at the University of Missouri, MaizeGDB has developed MaizeMine, a working interoperability data warehouse based on the Intermine software package. MaizeMine accelerates genomic analysis by enabling researchers without scripting skills to create and export customized annotation datasets merged with their own research data for use in downstream analyses. MaizeMine integrates genomic sequences and gene annotations from the B73_RefGen_v3 and v4 genome assemblies, Gene Ontology, protein annotations (UniProt), protein families and domains (InterPro), homologs (Ensembl Compara), and pathways (CornCyc, KEGG, Plant Reactome). It also provides simple and sophisticated search tools, including a keyword search, built-in template queries with intuitive search menus, and a QueryBuilder tool for creating custom queries. The Regions search tool executes queries based on lists of genome coordinates, and supports both B73_RefGen_v3 and v4. The List tool allows users to upload identifiers to create custom datasets, perform list operations such as unions and intersections, and execute template queries with lists. When used with gene identifiers, the List tool automatically provides gene set enrichment for GO and pathways, with a choice of statistical parameters and background gene sets.
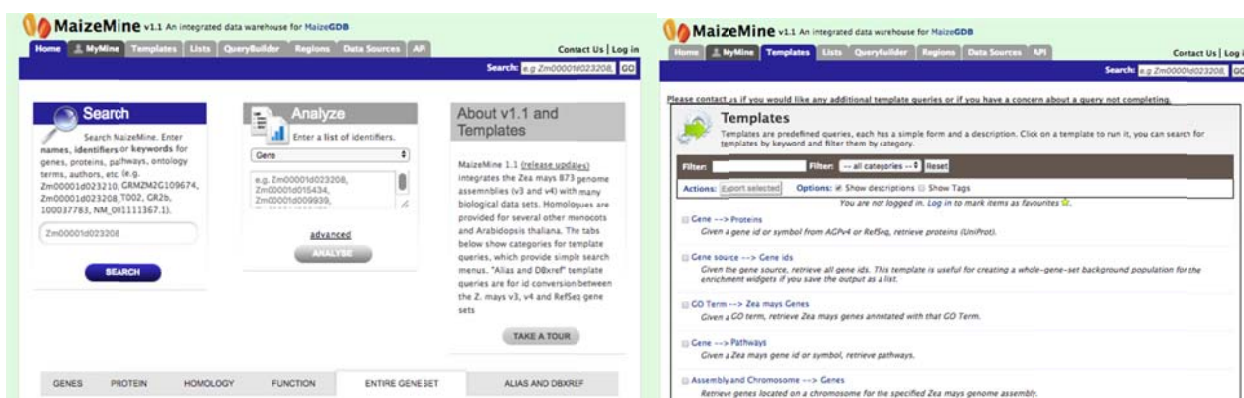

MaizeMine is a user-intuitive data warehouse based on the B73_RefGen_v3 and v4 assemblies that allows users to create and export custom lists in variety of formats. It contains over 20 pre-built templates that allows users to retrieve data in the most frequently requested data configurations.

Data Highlights for 2017
- CornCyc 8.0 was released at MaizeGDB for the v4 gene models after being developed by the Plant Metabolic Network in collaboration with MaizeGDB.
- The maize core bin markers have been converted to SNP-based markers based on pan genome anchors. This will allow easy projection of core bin SNPs on to the numerous new whole genome assemblies being hosted at MaizeGDB.
- MaizeGDB has a new genome browser track for over 300,000 EMS point mutations and over 5,000 Indels contained in 1,942 sequence indexed maize lines that are available from the Chinese Academy of Agricultural Sciences.
- MaizeGDB curators, in collaboration with personnel at the NCBI have selected over 300 RNA-SEQ datasets from over 90 B73 maize tissues for inclusion in the pipeline used to develop the NCBI RefSeq gene models for B73_RefGen_v4.
- MaizeGDB has created a new genome browser track for B73_RefGen_v3 and v4 for NCBI dbSNP's 58 million non-redundant SNPs with permanent RS#s. This is critical, as these non-redundant RS#s represent the best way for tracking SNPs across new whole genome assemblies.
- Student curators, guided by MaizeGDB curators, are systematically curating Maize Meeting abstracts to identify and document information on gene model function that would otherwise be missed by automated annotation algorithms.

- MaizeGDB in collaboration with the Schnable laboratory at ISU has uploaded over 1200 gel pictures for a survey of 24 inbred lines with ISU molecular markers. Information on molecular markers is one the most frequently requested bulk downloads at MaizeGDB.
- MaizeGDB has developed MaizeDIG which is a curation platform that allows phenotype images to be displayed on a genome browser enabling MaizeGDB's users to view the phenotypic images within its genomic context. See Figure below.
- MaizeGDB worked with collaborators from the Plant Breeding, Technical University of Munich, Germany to make publically available two whole genome sequences of the maize Flint lines EP1 and F7. The European Flint reference sequences provide additional diversity to other previously sequenced lines and complement the maize pan-genome.
- MaizeGDB worked with NRGene to submit four whole genome assemblies (EP1, F7, PH207, W22) to GenBank.
- MaizeGDB is working with collaborators at NRGene and the Chinese Agricultural University in Beijing, China to make the Mo17 genome sequence publically available.
- MaizeGDB worked with collaborators at the Iowa State University and the ISU Transformation Facility to sequence, assemble, and annotate the maize line B104. New data includes 12X PacBio sequences. B104 is the inbred line used by the Transformation Facility to produce all of its primary transformants.
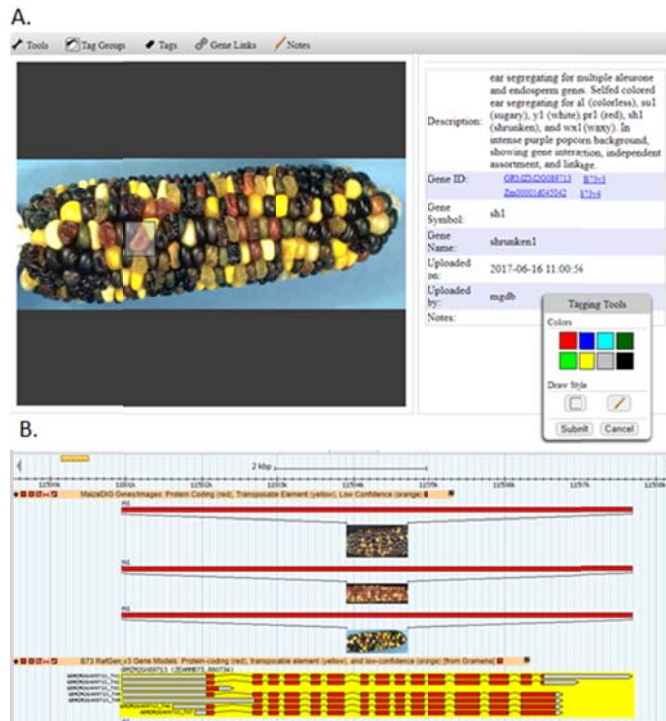
MaizeDIG is a web-based tool that allows curators to manually annotate genotypic-phenotypic relationship on images. MaizeDIG has been preloaded with over 1,000 mutant phenotypes images that have already been associated with a maize gene/gene model. These images are then dynamically made available as a custom track on the MaizeGDB Genome Browser. Panel A shows the curation interface for mutant phenotype image associated with the sh1 gene. Panel B shows the sh1 region on the MaizeGDB Genome Browser with the MaizeDIG and gene model tracks.

MaizeGDB Outreach Activities

MaizeGDB organized a series of six workshops at the 59th Annual Maize Genetics Conference in St. Louis, Missouri (https://community.maizegdb.org/workshops/workshops.php). MaizeGDB team members Maggie Woodhouse, Jesse Walsh, and Lisa Harper led four of the workshops that were filled to capacity and standing room only. Curator Lisa Harper also organizes monthly seminars for the AgBio working group, which cover topics on biocuration and biological databases. Lisa was also one of the organizers for an April 2017 meeting of AgBio Data working group in Salt Lake City, Utah. Over 45 scientists from 21 different institutions representing 32 agriculturally relevant databases and resources attended. This meeting covered topics in curation, metadata and persistence, database storage, data sharing using web services, policy, and communication. From these discussions and previous work, the AgBio working group is currently writing a white paper on best practices in each of these areas. This will help enhance genetic, genomics and breeding research outcomes through standardization of practices and protocols across agricultural databases. MaizeGDB, in collaboration with Gramene (www.gramene.org), also organizes the Agricultural Database booth at the 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. In 2017, over 20 databases participated in the Agricultural Database booth.

Community Leadership Roles
- USDA-ARS Executive Scientific Advisory Committee – C. Andorf (new)
- USDA-ARS CERES High Performance Computing Policy Committee – J. Portwood (new)
- USDA-ARS Database Committee - J. Portwood (new)
- 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 (new)

MaizeGDB also provides technical support for the Maize Genetics Executive Committee, Maize Genetics Conference Steering Committee, the McClintock Prize for Plant Genetics and Genome Studies, and the Maize Community Awards.

Staffing at MaizeGDB

Staffing at MaizeGDB has been relatively stable over the past year with a few notable changes. The biggest change at MaizeGDB is that Maggie Woodhouse, previously in a part time, temporary position, has moved to a full time permanent position and relocated from Sacramento to Ames, IA. Maggie's computational skills in genome biology have been invaluable in recruitment, analysis, and display of whole genome assemblies. John Portwood has moved from his temporary position to a permanent position ensuring his continued presence at MaizeGDB. Carson Andorf continues to serve as the lead scientist for MaizeGDB. Lisa Harper (0.5 FTE) and Jack Gardiner (1.0 FTE) are curators. Retired curator Mary Schaeffer continues to work part time (0.2 FTE) at MaizeGDB. Ethy Cannon works half-time on the project as a bioinformatics engineer. She also holds a half-time scientist position with the Computer Science Department at Iowa State. Jesse Walsh is a postdoctoral fellow. Current students funded though MaizeGDB are Nancy Manchanda (B014 assembly), Kyoung Tak Cho (MaizeDIG), David Schott (tool development), and Mandy Dietze (curation). There are still two vacant positions at MaizeGDB: Computational Biologist (Scientist) and Scientific Programmer (IT-Specialist).

Publications for 2017

Sen T, Braun B, Schott D, Portwood J, Schaeffer M, Harper L, Gardiner J, Cannon E, Andorf C. Surveying the Maize Community for their Diversity and Pedigree Visualization Needs to Prioritize Tool Development and Curation. Database (Oxford). 2017 Jan 1;2017. doi: 10.1093/database/bax031.

Anderson SN, Johnson CS, Chesnut J, Jones DS, Khanday I, **Woodhouse M**, Li C, Conrad LJ, Russell SD, Sundaresan V. The Zygotic Transition Is Initiated in Unicellular Plant Zygotes with Asymmetric Activation of Parental Genomes. Dev Cell. 2017 Nov 6;43(3):349-358.e4. doi: 10.1016/j.devcel.2017.10.005.

Gage JL, Jarquin D, Romay C, Lorenz A, Buckler ES, Kaeppler S, Alkhalifah N, Bohn M, Campbell DA, Edwards J, Ertl D, Flint-Garcia S, **Gardiner J**, Good B, Hirsch CN, Holland J, Hooker DC, Knoll J, Kolkman J, Kruger G, Lauter N, Lawrence-Dill CJ, Lee E, Lynch J, Murray SC, Nelson R, Petzoldt J, Rocheford T, Schnable J, Schnable PS, Scully B, Smith M, Springer NM, Srinivasan S, Walton R, Weldekidan T, Wisser RJ, Xu W, Yu J, de Leon N. The Effect of Artificial Selection on Phenotypic Plasticity in Maize. Nat Commun. 2017 Nov 7;8(1):1348. doi: 10.1038/s41467-017-01450-2.

Maher KA, Bajic M, Kajala K, Reynoso M, Pauluzzi G, West DA, Zumstein K, **Woodhouse M**, Bubb K, Dorrity M, Queitsch C, Bailey-Serres J, Sinha N, Brady SM, Deal RB. Profiling of accessible chromatin regions across multiple plant species and cell types reveals common gene regulatory principles and new control modules. Accepted to Plant Cell.

Wimalanathan K, Friedberg I, **Andorf C**, Lawrence-Dill C. Maize GO Annotation - Methods, Evaluation, and Review (maize-GAMER). bioRxiv 222836; doi: https://doi.org/10.1101/222836 (submitted to Frontiers in Plant Science).

Schott DA, Vinnakota AG, Portwood JL, **Andorf CM**, Sen TZ. SNPversity: A Web-based Tool for Visualizing Diversity. Database (Submitted).

Acknowledgements.