



December 2014

### Site Redesign:

The MaizeGDB team has been working on a multi-year effort to redesign the MaizeGDB interface. We announce the release of the Beta version (<http://beta.maizegdb.org>) of the new MaizeGDB interface on December 5th. This is the final major release before we transition to the new website. We plan to replace the MaizeGDB website with the new interface sometime prior to the Maize Meeting in March 2015. In addition to the new look and feel and improved navigation, improvements include hardware upgrades, access to assembly data sets, and the release of a new and faster genome browser.

### Genome Assembly Stewardship:

In addition to representing the latest version of the B73 assembly, MaizeGDB has undertaken proactive stewardship. As described in the December 2014 Report, this is a collaboration with the NCBI “Genome Reference Consortium”, the researchers engaged in the primary assembly, and the maize research community. We have submitted updates for 5,000 BACs to GenBank to better reflect the order and orientation between the BACs and the reference assembly. We have begun collecting issues from the community, using a tracker that aligns well with the GRC tools for representing “patches” or pending updates. For more information see GRC, Church DM et al (2011) PLoS Biol 9:e1001091, <http://www.ncbi.nlm.nih.gov/projects/genome/assembly/grc/>.

### Genome Browser:

The latest assembly version, B73 RefGen\_v3 has been released with two updated annotations from the Maize Sequencing Project and the NCBI pipeline. Core bin markers have been added as a track, along with new data referred to below. The B73 RefGen\_v3 browser has a total of 28 tracks and the RefGen\_v2 browser has 56 tracks.

### New data highlights:

- “Guide to Maize Mutant Phenotypes” (<http://mutants.maizegdb.org/doku.php>) has been released at MaizeGDB by MG Neuffer and associates. All images are part of a larger image collection at MaizeGDB, but only highest quality images were selected for this work, along with updated description.
- UniformMu insertion data now includes 57,582 insertions, represented in 9408 seed stocks accessible from the Maize Genetics Cooperation Stock Center (MGCSC), and accessible from the genome browser.
- Ds insertions carrying a GFP protein tag represented by 1714 seed stocks from Hugo Dooner, many available at the MGCSC. These have been added to the

Dooner Ac/Ds track on the genome browser. Li, Y et al. 2013. *Methods in Molecular Biology* 1057:83-99

- Means values phenotype scores for IBM and NAM populations from literature (8 papers, some 60 traits), integrated at MaizeGDB with related information about how traits were measured, location(s), special conditions (e.g. day length), and descriptors (ontology terms) applicable to all crops, to aid in interoperability with data from other crops. Interface under development needs to accommodate these data types.
- CORNFED consensus genetic maps, based on European double haploid mapping populations and Illumina MaizeSNP50 array. Giraud, H et al. 2014. *Genetics* 114.169367. All loci are also on a genome browser track.
- Quantitative proteomics data for eight seed tissues with more expected in near future from the UCSD maize proteomics project (<http://maizeproteome.ucsd.edu/>). These are accessible on the genome browser. Walley JW et al 2013 *Proc Natl Acad Sci* 110:E4808-17.
- RNA-seq data from S Kaeppler group, release pending publication. These data overlap the tissue set earlier characterized in the Nimblegen microarray data.

#### Staff changes:

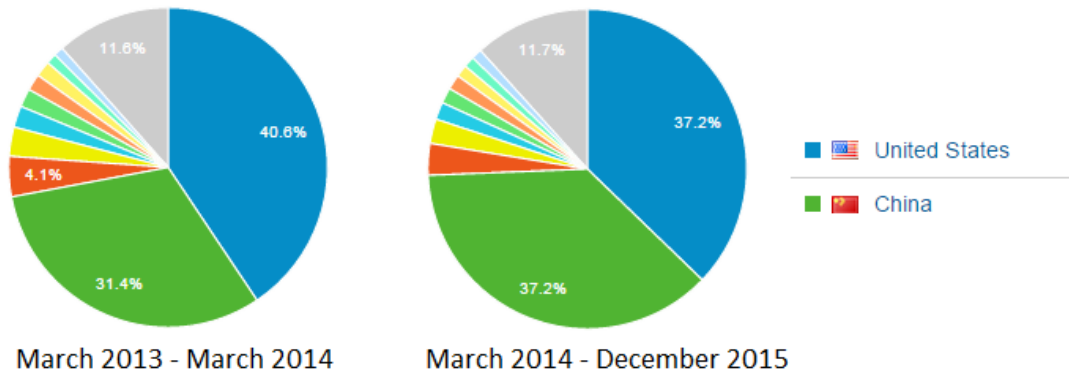
New funding has been supplied to MaizeGDB, as part of the Big Data initiative to “facilitate the use of genomic and genetic data, information, and tools for germplasm improvement, thus empowering ARS scientists and partners to use a new generation of computational tools and resources” (from MaizeGDB’s amended 5-year project plan). Part of this funding is being used to support undergraduate and graduate students. Miranda Leek (undergraduate), is aiding in issues relating to Genome Assembly Stewardship, and Gene Ontology annotation of papers recommended by the Editorial Board, and providing Facebook and Twitter feeds. Vivek Sribalusu (Master’s student in Computer Science) is implementing a maize instance of InterMine (MaizeMine), to allow flexible, user-specified queries to underlying datasets at MaizeGDB. Abhinav Vinnakota (Master’s student in Computer Science) is developing a database infrastructure to allow web service access to large scale maize diversity data. The additional funds were used to rehire three former MaizeGDB employees, Ethy Cannon (Bioinformatics Engineer), Bremen Braun (Interface developer), and Jackie Richter (Bioinformatics graduate student). We are also in the process of converting John Portwood from a part-time student programmer to a full-time scientific programmer. Other news: Carolyn Lawrence and Darwin Campbell have taken positions with Iowa State University, but remain close collaborators to MaizeGDB and are available for consulting and training. We anticipate filling those positions in the near future. Jack Gardiner shares his time with the Lawrence lab and with MaizeGDB.

#### Outreach:

Tutorials were provided by Lisa Harper at the 56th Annual Maize Genetics Conference hosted in Beijing, China. We also created Chinese versions of the MaizeGDB homepage and the 2014 Maize Meeting website. After the 2014 Maize Meeting, the usage statistics from China increased by over 10%, making domestic and Chinese usage nearly equivalent over the same time period (see figure below). May 2014 Jack Gardiner and the NCGA produced a podcast about the shoot apical meristem. For more information see the home page at

MaizeGDB. MaizeGDB continues to provide support for the Maize Genetics Conference (preparation of abstract booklet and program and maintaining the conference website), and this year a survey from the Maize Genetics Executive Committee on research directions and infrastructure priorities with results posted at MaizeGDB.

MaizeGDB Usage Statistics before and after Beijing Maize Meeting



Publications:

**Andorf CM**, Kopylov MS, Dobbs D, Koch KE, Stroupe ME, **Lawrence CJ**, and Bass HW. G-quadruplex motifs in maize (*Zea mays* L.) reside at specific sites in thousands of genes coupled to energy stress pathways, including hypoxia, low sugar, and nutrient deprivation. *JGG*. 2014 Dec. 10.1016/j.jgg.2014.10.004

Campbell MS, Law M, Holt C, Stein JC, Moghe GD, Hufnagel DE, Lei J, Achawanantakun R, Jiao D, **Lawrence CJ**, Ware D, Shiu SH, Childs KL, Sun Y, Jiang N, Yandell M. 2014. MAKER-P: a tool kit for the rapid creation, management, and quality control of plant genome annotations. *Plant Physiol*. 2014. 164:513-24

Law M, Childs KL, Campbell MS, Stein JC, Holt C, Panchy N, Lei J, Jao D, **Andorf CM**, **Lawrence CJ** Iowa State, Ware D, Shiu SH, Sun Y, Jiang N, Yandell M. 2014. Automated update, revision and quality control of the *Zea mays* genome annotations using MAKER-P improves the B73 RefGen\_v3 gene models and identifies new genes. *Plant Physiol*. 2014 Nov 10. pii: pp.114.245027.

Mao Y, Van Auken K, Li D, Arighi CN, McQuilton P, Hayman GT, Tweedie S, **Schaeffer ML**, Laulederkind SJ, Wang SJ, Gobeill J, Ruch P, Luu AT, Kim JJ, Chiang JH, Chen YD, Yang CJ, Liu H, Zhu D, Li Y, Yu H, Emadzadeh E, Gonzalez G, Chen JM, Dai HJ, Lu Z. 2014 Overview of the gene ontology task at BioCreative IV. *Database (Oxford)* 2014: 10.1093/database/bau086.

**Schaeffer ML**, **Sen TZ**, **Lawrence CJ** 2014 Databases. IN *Genetics, genomics and breeding of maize*. Editors Wusirikia R, Bohn M, Lai J, Kole C, CRC Press Boca Raton FL, pp 215-235.

Van Auken K, **Schaeffer ML**, McQuilton P, Laulederkind SJ, Li D, Wang SJ, Hayman GT, Tweedie S, Arighi CN, Done J, Müller HM, Sternberg PW, Mao Y, Wei CH, Lu Z. 2014. BC4GO: a full-text corpus for the BioCreative IV GO task. Database (Oxford). 2014 Jul 28;2014. pii: bau074

Walsh JR, **Sen TZ**, Dickerson JA BMC Syst Biol. 2014 A computational platform to maintain and migrate manual functional annotations for BioCyc databases. Oct 12;8(1):115. [Epub ahead of print]

#### Acknowledgements:

Guidance is generously provided by the MaizeGDB Working group: A. Phillippy (Chair), A Barkan, Q Dong, D Jackson, T Lubberstedt, E Lyons, , M Sachs (*ex officio*), M Settles, and N Springer; the Maize Genetics Executive Committee: CJ Lawrence (Chair), E Buckler, M Timmermans, N Springer, J Birchler, S Hake, P Chomet, S Flint-Garcia, S Kaeppler, P Schnable; the Maize Nomenclature Committee: M Sachs (Chair), T Brutnell, H Dooner, C Du, T Kellogg, C Lawrence, M Schaeffer and P Stinard; and the MaizeGDB Editorial Board (2014): M Bartlett, M Hufford, A Gallovotti, C Hirsch, M Salazar.

We thank the USDA-ARS, the NSF, and the NCGA for funding.

Submitted by Carson Andorf, Bremen Braun, Ethalinda Cannon, Jack Gardiner, Lisa Harper, John Portwood, Mary Schaeffer, Taner Sen