V. MAIZE GENOME DATABASE

www.maizegdb.org

Over the course of 2009-10 we focused on implementing and improving tools and interfaces to represent the B73 reference genome assembly, which was published in November 2009 under the name B73 RefGen_v1 (Schnable et al., Science 326:1112-1115, 2009). Our efforts necessarily included the careful integration of genetic data with sequence. The centerpiece of our transition to represent the maize genome from a sequence-centric paradigm is the deployment of the MaizeGDB Genome Browser (Sen et al., Database (Oxford) 2009:bap020, 2009), which is available online at http://gbrowse.maizegdb.org and from links throughout the MaizeGDB website. In November 2009, two versions of the browser were available: a B73 BAC-based, pre-pseudomolecule view and the B73 RefGen_v1 browser. We encouraged maize researchers to align their data to the pseudomolecule, and provide data to MaizeGDB, so as to coincide with the formal release of the B73 RefGen_v1 assembly by the Maize Genome Sequencing Project (maizesequence.org) with the publication of their paper in November of 2009. Tracks on the B73 RefGen_v1 browser included: centromere regions and anit-CENH3 ChiP data from the Jiang/Presting groups; MAGIs aligned by the Schnable lab; ESTs, cDNAs and Ac/Ds insertions (of Vollbrecht/Brutnell) aligned by PlantGDB; and MIPS repeats and filtered gene sets from MaizeSequence.org. The quality of the filtered gene set from maizesequence.org is now shown as a color-coded track, served via DAS from PlantGDB, as is a track showing community annotations of gene structure. Over 2010, we added other community tracks: IBM 2008 Neighbors (Arizona Genome Institute AGI), ISU integrated IBM markers from the Schnable lab, UniformMu, and PLEXdb microarray probes (via PlantGDB). A third version, the B73 RefGen_v2, was released in March of 2010 by the Arizona Genomics Institute (AGI). RefGen_v2 was available via the MaizeGDB Genome Browser in May of 2010 and is the current default view of the genome at MaizeGDB. Tutorial videos outlining how to use the MaizeGDB Genome Browser and other items at MaizeGDB are available from the front page via links to “MaizeGDB Tutorials” and directly at http://outreach.maizegdb.org/.

The Locus Lookup Tool is key tool for integrating genetic data with the B73 reference genome sequence (Andorf et al. 2010), and has been upgraded to allow for more flexible searching. This tool finds pseudomolecule regions for genetically mapped loci, or pairs of loci, which may not yet be annotated to coordinates on the pseudomolecule. The tool works by (1) checking physical map coordinates to find out whether the locus is already placed. If so, the physically mapped locus is highlighted in red in the region returned within the browser window. If not, the tool (2) checks the locus record at MaizeGDB to find out if any BACs are known to detect the locus and that BAC is returned within its genomic context. If not, (3) genetically mapped probes that are nearest the input locus are identified, the tool checks whether those probes have known genomic coordinates (working outward until appropriate probes are identified) and finally the region of the genome contained by the identified probes is reported with bounding probes shown in red. The tool is an integral part of the MaizeGDB Genome Browser, and can be implemented for other GBrowse instances deployed at other sites and supporting data for other species. In addition we have implemented updates to several other interfaces including: BLAST, cDNA, EST, BAC, and Mo17 SNP records. The Mo17 SNP pages include a detailed view of sequence centered on a specified Mo17 SNP. This region shows any nearby insertions, deletions, or substitutions and the user can change the region size to 61, 241,401, or 561 base pairs as needed. An integrated tool also ‘predicts’ the Mo17 sequence for the specified region.

The Popcorn resource (http://popcorn.maizegdb.org) has continued to develop. We provide access to 104 projects and 132 resources, searchable by category (including, e.g., sequencing, mapping, mutation, bioinformatics, and breeding). Listings are regularly updated by MaizeGDB and POPcorn staff. In addition, project PI’s can submit new projects and to correct project information directly. Participating projects that support BLAST are searchable by two utilities: POPcorn BLAST, which permits researchers to BLAST against maize sequence at multiple sites, and the Sequence-Indexed Data Search, which carries out multi-step searches of sequence-indexed data. Prototype versions of POPcorn BLAST and the Sequence-Indexed Data Search were released in February of 2010. Sites available to POPcorn BLAST include: GRASSIUS, MAGIs, MaizeGDB, PalomeroTolulqueño.org, PlantGDB, PlexDB, and NCBI. Examples of queries supported by the Sequence-Indexed Data Search include mutant seed stocks from the Ac/Ds. PML, Mu insertions, TILLING, and UniformMu projects as well as more complex queries, such as phenotypes associated with sequences similar to the query. At this time, the latter query returns phenotypes, and various alleles/variations associated with each phenotype, often from more than one gene. To document provenance of the data and avoid poor interpretation of the results, POPcorn collects attribution and citation information, links to collaborators’ sites, and displays short descriptions prominently on all POPcorn pages where collaborator data and project descriptions are made available.

Our focus on data curation has been reinvigorated with: (1) further development of the ‘maize gene review’, first released in 2009 as part of the MNL, and accessible online at http://www.maizegenereview.org; (2) the addition of two new personnel – Lou Butler, who entered much of the original data and images from the M. Gerry Neuffer collection into MaizeDB and assisted MGN in the preparation of Mutants of Maize, and Jack Gardiner to integrate CIMMYT data, as well as GRASSIUS and UniformMu resources. The quality of existing mutant images is being assessed by MGN and LB, and where image quality is determined to be inferior, new images are being generated. We are using Jack’s expertise with the Maize Microarray Project (Arizona) to begin integrating sequence-based gene expression data types, including both microarray and RNA_seq, and others as they are made available. Literature curation is largely based upon recommendations from the Editorial Board input, which averages 5 articles/month, with additional inputs from community curator, Ed Coe and requests from the community. We have just begun a new collaboration with Truman State University to provide Gene Ontology (GO) annotation for interesting genes as revealed by the SAM (shoot apical meristem) Project (see also http://sam2.truman.edu/index.html for more information).
Outreach has been greatly enhanced by several new tutorials on both the use of MaizeGDB, and the understanding of new datatypes for maize. The list of tutorials includes 4 focused on the B73 sequence: MaizeGDB Genome Browser; How was each B73 BAC sequenced; BAC based B73 Genome Assembly Cavets; and All about B73 RefGen v1. These are prominently featured on the homepage, next to our RSS feed and Facebook links. The tutorial page also provides links to other tutorials of interest to maize researchers. We continue to provide on-site personal tutorials on request by locations with several maize research groups.

We have made a major upgrade to our infrastructure by implementing a virtual server environment. This aligns with Executive Order 13423: Strengthening Federal Environmental, Energy and Transportation Management, sec. 2(h); sec.3(a and f) by improving energy efficiencies and extending computer replacement lifecycles, by reducing the number of physical computers, while maximizing existing hardware capacity. The virtual servers interact with each other and users just like physical servers, without dedicating new of specified hardware to a specific purpose. New virtual servers allow the testing of new operating systems, software patches and application interactions with the ability to roll changes out of the testing procedure without starting over. The virtual environment is made up of three identical servers; each acts as host and can pick up services provided by a failing host, with priority given to production virtual servers.

Noteworthy: Lawrence, Sen, Andorf and Campbell were members of the group awarded the USDA-ARS Midwest Area’s Outreach, Diversity and Equal Opportunity Award in 2009 for Outreach to American Indians (led by C. Lawrence and C Gardner; funded by the NSF). Campbell was awarded the USDA-ARS Midwest Area’s Research Support Award in 2009.

Publications 2009, 2010

Sen, TZ; Andorf, CM; Schaeffer, ML; Harper, LC; Sparks, ME; Duvick, J; Brendel, VP; Cannon, E; Campbell, DA; Lawrence, CJ (2009) MaizeGDB becomes ‘sequence-centric’. Database (Oxford) 2009:bap020.

Sen, TZ; Harper, LC; Schaeffer, ML; Andorf, CM; Seigfried, TE; Campbell, DA; Lawrence, CJ (2010) Choosing a genome browser for a Model Organism Database: surveying the maize community. Database (Oxford) 2010:baq007


Acknowledgements

Guidance is generously provided by the MaizeGDB Working Group: M. Pop (Chair), A. Barkan, V. Brendel, M. Freeling, O. Hoekenga, A. Lamblin, T. Lubberstedt, E. Lyons, K. McGinnis, L. Mueller, M. Sachs, P. Schnable, A. Sylvester, and D. Ware); the Maize Genetics Executive Committee; T. Brutnell (Chair), P. Schnable, M. Sachs, V. Walbot, W. Tracy, S. Wessler, J. Bennetzen, B. Boston, E. Buckler, and C. Lawrence; the Maize Nomenclature Committee; H. Dooner, (Chair), T. Brutnell, V. Chandler, C. Hannah, T. Kellogg, M. Sachs, M. Scanlon, M. Schaeffer, and P. Stinard, the MaizeGDB Editorial Board (2009, 2010) M Arteaga-Vazquez, P Brown, J. Dorweiler, P Balint-Kurti, N Lauter, P McSteen, E Vollbrecht, R Wisser, and C Weil. Funding for MaizeGDB has been provided by the USDA-ARS with additional support from NSF (0743804, 0701405, 0703273, and 0965380) and via CIMMYT from USAID.

Submitted by the MaizeGDB Team: Schaeffer, ML; Andorf, CM; Birkett, S; Braun, B; Campbell, DA; Cannon, E; Gardiner, J; Harper, LS; Sen, TZ; and Lawrence, CJ