

## MaizeGDB – Past, Present, and Future

Carolyn J. Lawrence

Corn Insects and Crop Genetics Research Unit, USDA-ARS  
Department of Genetics, Development, and Cell Biology, Iowa State University, Ames, IA 50011, USA  
E-mail: carolyn.lawrence@ars.usda.gov



The Maize Database turns twenty this year. Many of us trace our academic roots back to Emerson: we can do likewise for MaizeGDB. In 1928 Emerson held a get-together of maize researchers in his hotel room at the American Association for the Advancement of Science meetings. It was this “Cornfab” that marked the beginning of

the Maize Genetics Cooperation (see [Figure 1](#); reviewed in detail in [Kass et al, 2005](#)). Shortly thereafter, Emerson and Beadle compiled the first letter to all Cooperators, later referred to the Maize Newsletter (MNL), volume 1. Note that the MNL continues to be published to this day. In 1932 Marcus Rhoades took on the role of Secretary for the Cooperation and founded the Stock Center. In 1956, the first maize-focused peer-reviewed journal appeared: *Maydica*.

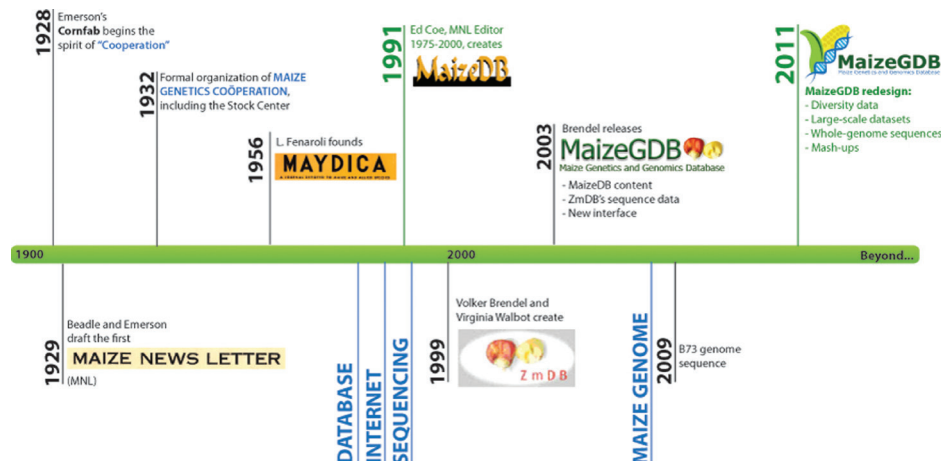
As information on maize as a crop and a model organism accrued, so did methodologies for storing and accessing data. In the 1970’s, Edgar Codd at IBM developed the theory that resulted in relational database technologies ([Codd, 1970](#)), almost simultaneously, ARPANET (forerunner to the modern internet; [Salus, 1995](#)) came online. With the invention of PCR in 1983 driving the ability to amplify then sequence DNA ([Bartlett and Stirling, 2003](#)), the pieces were all in place for generating lots of maize genome data, storing it relationally, and making the data available online. This was apparent to MNL editor Ed Coe who responded by creating MaizeDB, the original Maize Database in 1991 ([Polacco and Coe, 1999](#)). Initially available via ARPANET, MaizeDB was one of the first biological databases online, predating even GenBank (which came online in 1992; [Benson et al, 1997](#)). In 1999 Volker Brendel and Virginia Walbot created ZmDB, a maize sequence database ([Dong et al, 2003](#)), and in 2003 MaizeDB and ZmDB were combined into what is now called MaizeGDB ([Lawrence et al, 2004](#)).

Currently MaizeGDB is accessed more than ever before with much of the increased activity occurring as a result of a shift to a more sequence-centric representation of data including the addition of the B73 reference genome sequence to the MaizeGDB re-

source ([Sen et al, 2009](#)). At the same time, researchers continually articulate a desire for a more versatile and interactive experience that is in keeping with the functionality they have become accustomed to online at commercial websites such as Amazon and Facebook. As we consider how best to address these needs, the scale of sequence data anticipated to be coming our way to support diversity analyses seems to grow geometrically and solutions for how best to offer better access to what is now being referred to as “Big Data” are few and far between (see *Nature* vol. 455 no. 7209: a special issue dedicated to this problem).

With these challenges in mind, the MaizeGDB Team has resolved to meet researchers’ articulated needs by completely remaking MaizeGDB over the course of the next two years. Although this timeline is fairly ambitious, I assure you that the current staff (see <http://www.maizegdb.org/personnel.php>) has the moxie and expertise required to pull it off! The redesign focuses not only on updating the look and feel of the Web interface, but also on making use of recently developed tools and technologies that underlie the site’s behavior to improve load times and responsiveness. Our goal is to gather and make use of this and other technologies to improve your access to data via MaizeGDB.

A key aspect of the planned site redesign is incorporation of information stored off-site into what is displayed via the MaizeGDB website. Mixing information from disparate resources for a single view using Web technologies results in what is called a “mash-up” ([Clarkin and Holmes, 2007](#)). You may have seen such representations at Yahoo, Google, or even Facebook where your profile information is used to “mash-up” ads and other items tailored to appeal to your tastes alongside other information you provided. Here’s an example of how this could help with integrating maize resources: if you were browsing information related to the *bronze1* locus and a stock were noted in the GRIN (<http://www.ars-grin.gov/npgs/>) database as having a particular allele of *bz1* as well as an image of bulk kernels available, that image could be shown at MaizeGDB even though the image file itself would be stored at GRIN rather than on the MaizeGDB servers. Likewise, information made available from any participating off-site resource, e.g., GenBank ([Benson et al, 2011](#)), Gramene ([Youens-Clark et al, 2011](#)), the iPlant



**Figure 1** - Timeline leading up to MaizeGDB's upcoming redesign. From 1928 through today, the spirit of Cooperation has underscored maize research. In 1929 the first MNL was drafted, the Stock Center was founded in 1932, Maydica was established in 1956, and MaizeDB came online in 1991. In 1999 the first all-maize sequence database ZmDB was created, and in 2003 MaizeDB and ZmDB were combined to create MaizeGDB. In 2009 the draft sequence of inbred line B73 was published, and in 2011, MaizeGDB announced an endeavor to completely remake MaizeGDB from the ground up.

Collaborative (<http://www.iplantcollaborative.org/>) etc, could be made accessible via MaizeGDB even if the full depth of information has not been curated into the MaizeGDB database. Indeed this already is the case for some sequence-based datatypes served via the MaizeGDB Genome Browser including the "B73 RefGen\_v2 Gene Models: Quality" track provided by Volker Brendel's PlantGDB (Duvick et al, 2008) group (see Figure 2; to learn about how gene model quality is calculated, visit <http://plantgdb.org/GAEVAL/docs/integrity.html>). To get the information from PlantGDB to show up at MaizeGDB, a file is served by PlantGDB via a technology called DAS (the Distributed Annotation System; Dowell et al, 2001) that specifies how to represent the gene model graphically including its genomic location, intron/exon structure, and coloration. MaizeGDB's instance of GBrowse (Sen et al, 2010; Stein et al, 2002) has a built-in DAS client that allows it to consume the data served by PlantGDB. The Web interface at MaizeGDB displays the image of the gene model, and a figure legend is accessible by mouse-over or clicking the track title's box marked by a question mark. We can serve various types of data from off-site via MaizeGDB, and are in the planning stages now to determine how best to make use of data made available by other long-lived repositories as we remake MaizeGDB.

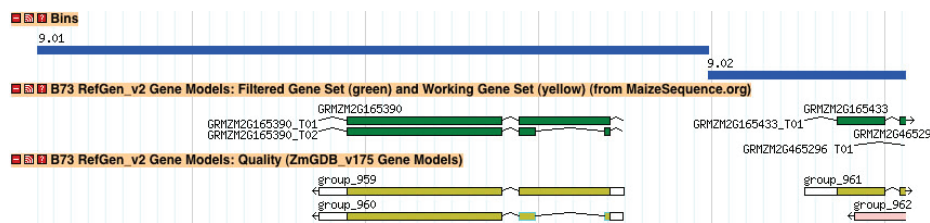
Given that a complete site redesign is planned, a researcher recently asked me whether a "Maize Breeder's Toolbox" would be deployed given that a few other crop websites have created such a view. I said no. Why not? Because researchers regularly cross-train students and interesting research questions often cross arbitrarily drawn boundaries with respect to discipline. Any sampling of tools tailored to meet the general needs of scientists in a given field will necessarily lack other tools useful to address particular research questions. Learning to use nothing

but a sample of the tools that would be selected to populate a toolbox would limit any researcher's capacity to make full use of the data and would therefore limit the use of and access to all appropriate and available data for a given research question.

The perceived divide between the disciplines of breeding and basic biology shrinks constantly, and the number of individuals who regularly work in both fields is on the increase. In the 1920's and 30's maize researchers well understood both breeding and basic biology and took a holistic approach to understanding maize as a crop, a species, and a model for research. Over time, the disciplines of basic biology and breeding, though interrelated, diverged. However, it seems that there is renewed appreciation for understanding all aspects of maize biology to better address problems. The number of us working to bridge the basic biology/breeding divide both in training students and in our own research is growing rapidly. In addition, the formation of collaborative teams that address real-world problems through basic scientific investigation typifies how we do science in the 21st century. For maize research, I am happy to report that cooperation remains in vogue: everything old is new again.

## Acknowledgements

Thanks to the MaizeGDB Team (Carson Andorf, Scott Birkett Bremen Braun, Darwin Campbell, Ethalinda Cannon, Jack Gardiner, Lisa Harper, Mary Schaeffer, and Taner Sen) for their excellent work and to the MaizeGDB Working Group (Alice Barkan, Owen Hoekenga, Anne-Francoise Lamblin, Thomas Lubberstedt, Eric Lyons, Karen McGinnis, Lukas Mueller, Mihai Pop Marty Sachs, Pat Schnable, and Anne Sylvester) and the community of maize researchers for their continued guidance and support. Thanks to Destri Andorf for her help with Figure 1, Jinsong Liu,



**Figure 2** - Data stored at MaizeGDB as well as offsite are available via the MaizeGDB Genome Browser directly. At the boundary between bins 9.01 and 9.02 is *bronze1*. Two transcript models for the *bz1* locus are predicted: GMZM2G165390\_T01 and T02. The upper track labeled “Bins” shows where the bin boundaries are predicted to lie, the next track shows the gene models. The “B73 RefGen\_v2 Gene Models: Quality (ZmGDB\_v175 Gene Models)” track shows that these two transcript models for *bz1* are quite good (exon background shading goes from: gray to gold with gold representing the best support). In this track, where exons conflict with expression evidence borders are shown as red (none in this example). Exons for which alternative splicing is indicated have a cyan border, and those with an alternative UTR are indicated with a pink background (see model to the right in this shot). To access color coding information from the context of the MaizeGDB Genome Browser directly, mouse over or click question mark boxes to the left of track names.

Jon Duvick, Shannon Schluetter, and Volker Brendel at PlantGDB for their work on the gene model quality assessment data and services. Work described here was supported by the USDA-ARS.

## References

- Bartlett J, Stirling D, 2003. A Short History of the Polymerase Chain Reaction. Springer
- Benson DA, Boguski MS, Lipman DJ, Ostell J, 1997. GenBank. Nucleic Acids Res 25: 1-6
- Benson, DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW, 2011. GenBank. Nucleic Acids Res 39: D32-37
- Clarkin L, Holmes J, 2007. Enterprise Mashups. The Architecture Journal 2007: October (<http://msdn.microsoft.com/en-us/library/bb906060.aspx>)
- Codd E, 1970. A relational model of data for large shared data banks. Communications of the ACM 13: 377-387
- Dong Q, Roy L, Freeling M, Walbot V, Brendel V, 2003. ZmDB, an integrated database for maize genome research. Nucleic Acids Res 31: 244-247
- Dowell RD, Jokerst RM, Day A, Eddy SR, Stein L, 2001. The distributed annotation system. BMC Bioinformatics 2: 7
- Duvick J, Fu A, Muppirala U, Sabharwal M, Wilkerson MD, Lawrence CJ, Lushbough C, Brendel V, 2008. PlantGDB: a resource for comparative plant genomics. Nucleic Acids Res 36: D959-965
- Kass LB, Bonneuil C, Coe EH Jr, 2005. Cornfests, cornfabs and cooperation: the origins and beginnings of the maize genetics cooperation news letter. Genetics 169: 1787-1797
- Lawrence CJ, Dong Q, Polacco ML, Seigfried TE, Brendel V, 2004. MaizeGDB, the community database for maize genetics and genomics. Nucleic Acids Res 32: D393-397
- Polacco M, Coe E, 1999. MaizeDB: the maize genome database. pp. 151-162. In: Bioinformatics: Databases and Systems. Letovsky S (ed). Kluwer, Norwell, MA
- Salus P, 1995. Casting the Net: From ARPANET to INTERNET and Beyond. Addison-Wesley Professional
- 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 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 2010: baq007
- Stein LD, Mungall C, Shu S, Caudy M, Mangone M, Day A, Nickerson E, Stajich JE, Harris TW, Arva A, Lewis S, 2002. The generic genome browser: a building block for a model organism system database. Genome Res 12: 1599-1610
- Youens-Clark K, Buckler E, Casstevens T, Chen C, Declerck G, Derwent P, Dharmawardhana P, Jaiswal P, Kersey J, Karthikeyan AS, Lu J, McCouch SR, Ren L, Spooner W, Stein JC, Thomason J, Wei S, Ware D, 2011. Gramene database in 2010: updates and extensions. Nucleic Acids Res 39: D1085-1094