

## V. MAIZE GENETICS AND GENOMICS DATABASE ([www.maizegdb.org](http://www.maizegdb.org))

### Big news.

- (1) MaizeGDB is reported as the fifth most accessed website by lead principal investigator of the National Plant Genome Initiative grants [National Research Council (USA), 2008. Achievements of the National Plant Genome Initiative and New Horizons in Plant Biology. Pp. 144-149, National Academies Press, Washington DC. <http://books.nap.edu/openbook.php?isbn=0309114187&page=145>].
- (2) USDA-ARS 5-Year Project Plan reviewed and accepted with no revisions required. Review process is described here: <http://www.ars.usda.gov/research/docs.htm?docid=1503>.
- (3) Genome Browser release and is continually updated.
- (4) Community genome annotation project funded for MaizeGDB and PlantGDB by USDA-ARS & NSF, respectively.

### New Staffing.

We welcome Carson Andorf, who joined the core team at Ames, IA in June of 2008. He fills the position vacated by Trent Seigfried in March of 2008. We also welcome Ethalinda (Ethy) Cannon to the group. Ethy is the Solution/Application Architect for the NSF-funded POPcorn project (described below).

Because the genomes of maize inbred lines B73 and Mo17 as well as the Palomero Toluqueño landrace population have been/are being sequenced, sequence data have taken center stage, and MaizeGDB's development has been refocused to allow a sequence-centric perspective linking genome sequences (especially B73) to existing genetic and physical map resources at MaizeGDB. Major endeavors related to creating the sequence-centric perspective include: incorporation of a genome browser into MaizeGDB to visually represent genome sequences; implementation of POPcorn (<http://www.maizegdb.org/POPcorn>), a portal ancillary to MaizeGDB that offers access to independent maize projects and will allow BLAST similarity searches of participating projects' datasets from a single point; and development of the infrastructure that enables the maize community to be involved in genome annotation.

### Genome Browser.

The initiative to implement the MaizeGDB Genome Browser was launched in early 2008, with the browser itself becoming available toward the end of 2008. During the planning stages, the MaizeGDB team considered a variety of genome browser applications and queried the maize community for their inputs. To gauge cooperators' impressions of existing software and to find out what functionalities they would like to have in a maize genome browser, a survey (accessible online at <http://www.maizegdb.org/blanksurvey.html>) was prepared and distributed on behalf of the Maize Genetics Executive Committee. A summary of the survey results is available online at [http://www.maizegdb.org/genome\\_browser\\_survey.php](http://www.maizegdb.org/genome_browser_survey.php) and a detailed description of the process is forthcoming (Sen et al., in preparation). Based upon results of the survey, GBrowse [Stein et al., 20002. The generic genome browser: a building block for a model organism system database. *Genome Res.* 12(10):1599-1610] was selected as the software for the MaizeGDB Genome Browser implementation.

The MaizeGDB Genome Browser was populated using data from: MaizeGDB, the Maize Genome Sequencing Consortium's MaizeSequence.org resource, PlantGDB [cDNAs, PUTs, *Ac/Ds* elements, etc.], Pat Schnable's group's MAGI resource, and the UniformMu Group's *Mu* insertions. Five volunteers (P. Balint-Kurti, S. Hake, D. Lisch, M. Muszynski, and V. Walbot) were selected to provide guidance on the MaizeGDB Genome Browser's development and ten were selected (A. Charcosset, O. Dugas, J. Estill, D. Hessel, D. Lisch, M. Muszynski, P. Scott, V. Walbot, R. Wang, and C. Alvarez-Meja) to serve as beta testers. Based upon their inputs and additional feedback from other researchers using the Genome Browser, custom tracks were developed including a 'BIN' track that divides the genome into 90 genetically defined segments familiar to most researchers and a 'Sequenced FPC contig' track that clearly delineates regions of the B73 genome that are not yet sequenced.

### Genome Browser Tools.

Guidance and Beta Tester volunteers (1) suggested the implementation of a tool to integrate genetic and genomic coordinates to allow researchers to identify regions of the genome where a genetically mapped locus may lie and (2) asked for mechanisms that would allow them to access genomic locations based upon sequence similarity analyses. Early in 2009, the Locus Lookup Tool was created and the BLAST tool was updated to meet these needs, respectively.

The **Locus Lookup** tool provides a way to find a mapped locus on the B73 sequence if its genomic coordinates are not known. Its algorithm uses the following steps: (1) if the locus has known physical coordinates, those coordinates are returned; (2) else, if there are there are probes that recognize the locus and the probes have known coordinates, then the coordinates of the BAC(s) that contain the probe(s) are returned; (3) else if the locus is on a given genetic map (IBM2 2008 Neighbors [default], NAM, Genetic 2005, or Genetic 2008), then the nearest left and right flanking marker with known physical coordinates are located and the coordinates of the BAC(s) that contain the region from the end position of the left flanking marker and the start position of the right flanking marker are returned; (4) else a

note that the region cannot be identified is returned. The strategy is diagrammed in Figure 1, and the GBrowse result for the *multiple archesporial cells1 (mac1)* locus is shown in Figure 2.

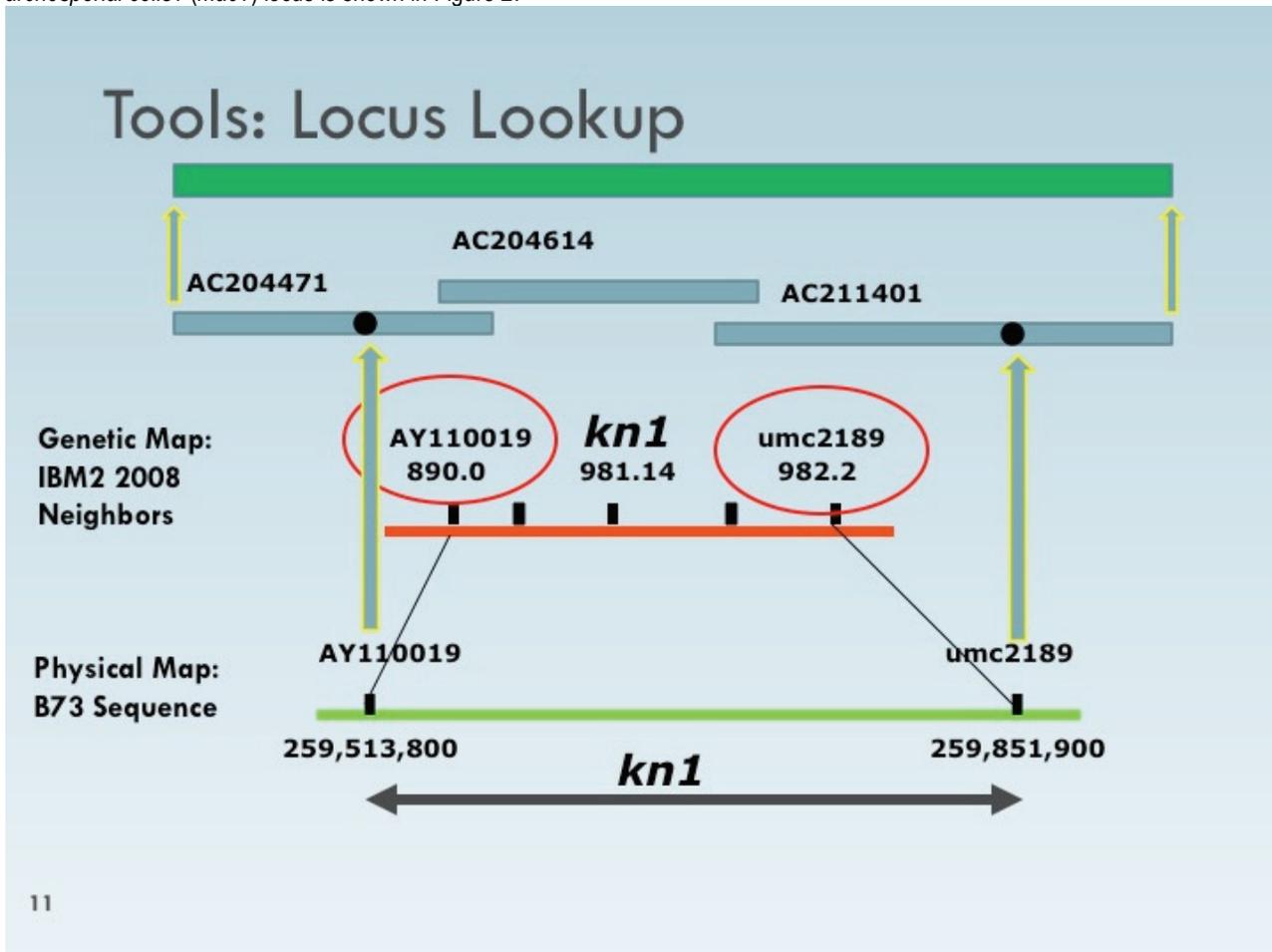


Figure 1. Strategy of Locus Lookup for the *kn1* *knotted1* locus.

MaizeGDB  jobs | upcoming events | sitemap  
 Maize Genetics and Genomics Database Useful Pages docs | bulk data | browse data | tools | login / register | links  
 home | Search all data for Go!

Locus Lookup found results on Chr 10

**Results for Chromosome 10 (Back to Top)**

**Chromosome coordinates based on MaizeSequence.org marker positions for: mac1**

The exact coordinates for mac1 were not found. The coordinates will be estimated based on the nearest flanking loci with known coordinates.

The Locus **mac1** is flanked by the following loci with known chromosome coordinates: (umc130 and umc1863)

The Locus **mac1** is **estimated** to be between position **11,284,700** and position **11,456,200** on **Chromosome 10** based on the map: **IBM2 2008 Neighbors 10**

This region is **171,500** base pairs long.

(Click [here](#) to go to the Genome Browser or scroll to the bottom of the page)

Locus	Chr	Map Position	Chr Start	Chr End
umc130	10	160	11,451,300	11,456,200
<b>mac1</b>	<b>10</b>	<b>165.21</b>	<b>0</b>	<b>0</b>
umc1863	10	165.52	11,284,700	11,289,600

Loci known to be between the flanking loci at map positions **160** and **165.52** on the **IBM2 2008 Neighbors map**:  
 ▶ [Click to expand details](#)

The Locus Lookup tool works by (1) checking physical map coordinates to find out whether the locus is already placed. If so, your physically mapped locus is highlighted in red in the region returned. 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.**

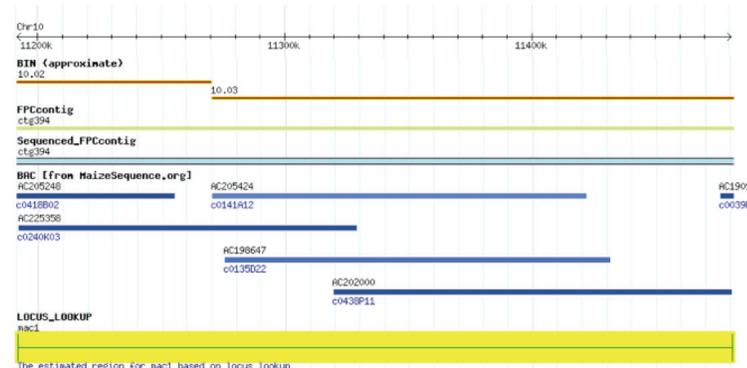
[More information](#)

**Your search for mac1 got a result with the logic described in (3) above.**

The Locus **mac1** is **estimated** between position **11,191,600** and position **11,480,700** on **Chromosome 10** located on one or more of the following BACs: AC225358 and AC205424 and AC198647 and AC202000

This region is **289,100** base pairs long.

(Click on image to go to the Genome Browser)



The estimated region for mac1 based on locus lookup

▶ [Expand Image Detail](#)

Try another term or map:

Search  for locus

Use only Genetic Map information

home | Search all data for enter terms here Go!

If you have some comments about this page, or about the site in general, fill out our easy to use [feedback form](#) without leaving this page! [Be sure to cite us!](#) 

Figure 2. The result page for the approximate genomic coordinates for the *multiple archesporial cells1* (*mac1*) locus when the Locus Lookup tool is used.

The MaizeGDB **BLAST tool** has been updated and integrated with the Genome Browser. If you BLAST at MaizeGDB, a thumbnail of the genomic region is returned and the hit itself can be loaded to the MaizeGDB Genome Browser to better visualize how those hits relate to other elements.

In April of 2009, links associated with the Locus Lookup Tool ranked first and sixth and links for BLAST ranked tenth and thirteenth among the top pages accessed at MaizeGDB. THANKS to all of you for directing our development!

### **New Data Highlights.**

Recombination-Based Maps. Some 260 individual NAM maps were added complete with map scores, provided by the “Molecular and Functional Diversity of the Maize Genome Project” (Diversity Project) pre-publication. Map names use the Diversity Project name, along with other specifications to help users. For example the map called “NAM Z002 B73xCML103 RI 1” permits lookups of all NAM maps by typing NAM in the search box; ‘Z002’ is the name used by the Diversity Project and the Maize Genetics Cooperation – Stock Center for the panel of stocks; “B73xCML103” indicates the parental lines, with B73 as maternal parent; “RI” indicates the panel type is recombinant inbred; “1” is the linkage group for the map. Alleles for each locus were included when the sequences had been submitted to GenBank and named based on the nucleotide position of the SNP used for mapping based on files prepared by the Diversity Project. The other major map set added was the ISU IBM INDEL Map7, which was provided pre-publication, with probe details and related GenBank accessions. Consensus maps have been updated to include: a new Genetic 2008 Map (based on the IBM2 map) to permit ready incorporation into the Neighbors computation; the IBM2 2008 Neighbors maps, with statistically defensible orders on contributing maps (IBM2 Frame) and with additional approximately ordered loci on the contributing maps (IBM2 Neighbors); and Bins maps now include the minimum tiling path BACs, using genomic coordinates provided by MaizeSequence.org.

UniformMu variations. with genomic coordinates defined by the UniformMu project were added, along with phenotypes of the stocks available from the MGCSC.

### **Outreach, tutorials, and movies.**

We plan on up to 3 trips each year to provide a 1-2 hour tutorial, tailored to specific questions or requests with one-to-one sessions after the tutorial. This year tutorials were held at Iowa State and UC Berkeley. Persons interested in scheduling a trip to their location should contact Lisa Harper or Carolyn Lawrence. Online PowerPoint presentations and movies that provide overviews of MaizeGDB, and specific topics are linked to the homepage. New movies describe the genome browser, and caveats in data interpretation. These caveats are expected to change as the sequence and its annotation are updated. We plan to update the movies as required. Cooperators are invited to request tutorials on a particular topic, using the link on the online page for this section.

### **Meeting Representation 2008 – early 2009.**

Sackler Workshop: NPGI’s planning meeting for the next 5 years; Biophysical Society Meeting; Genome Informatics Meeting; Plant and Animal Genome Conference (PAG); the Maize Genetics Conference; Corn Crop Germplasm Committee Meeting; MacWorld; 3<sup>rd</sup> International Biocurators Meeting. At PAG 2009, MaizeGDB hosted an exhibit booth for the Plant Genome DB Outreach Consortium that involved 13 database groups, 3 more than the previous year.

### **Editorial Board.**

We maintain an Editorial Board whose members contribute a paper each month to be highlighted at MaizeGDB. Data from these articles are manually curated and added to the MaizeGDB. The 2009 board includes Jane Dorweiler, Peter Balint-Kurti, Cliff Weil, and Randall Wisser. For more information see, [http://www.maizegdb.org/editorial\\_board.php](http://www.maizegdb.org/editorial_board.php), where you may now subscribe to an RSS feed.

### **Contributing your data to MaizeGDB.**

If you are developing a project that will generate large datasets and that you would like to submit to MaizeGDB, you need to contact Carolyn Lawrence before you submit the proposal. At this time we have subcontracts from two NSF-funded projects: (1) Construction of Comprehensive Sequence Transposon Resources for Maize with Don McCarty as PI ; and (2) The Grass Regulome Initiative: Integrating control of gene expression and agronomic traits across the grasses”, with Erich Grotewold as PI.

You may also contribute data directly. The easiest is very much like a ‘wiki’, where you simply add a comment using the annotation tool. You will first need to register, using the menu item ‘annotation’ on the top menu bar of the homepage. Once registered, every time you access MaizeGDB, you will be able to annotate. Annotations will appear in the monthly updates of the database. A second way is to use the community curation tools. Check the box when signing up for an account at MaizeGDB to get access to the curation tools, or ask for access directly using the feedback link at the bottom of any MaizeGDB page.

### **POPcorn.**

POPcorn (Project POrtal for corn), is a new website that addresses the difficulty maize researchers face when searching for and using differing data from multiple sources on the web. The first phase of POPcorn was released in April of 2009. This first release permits researchers to search for maize projects and web resources. The preliminary database contains 42 projects and 76 resources. POPcorn is continuously updated, so send along information on your projects and resources so that we can include them!

The second phase or POPcorn development will permit researchers to search sequences at multiple maize databases by submitting their query via the POPcorn interface. POPcorn will handle querying cooperator databases and returning the results in one place, freeing the researcher from having to make multiple queries at multiple sites.

The third phase will be mostly invisible to users, consisting of behind-the-scenes scripts, pipelines, and procedures for integrating project data into MaizeGDB when those projects come to end and need a permanent home for their data.

POPcorn is a two year, NSF-funded development project which will be folded into MaizeGDB upon completion. POPcorn can be found at <http://www.MaizeGDB.org/POPcorn>.

#### **Citing MaizeGDB.**

Most recent: Lawrence, CJ, Harper, LC, Schaeffer, ML, Sen, TZ, Seigfried, TE, Campbell, DA. (2008) MaizeGDB: The Maize Model Organism Database for Basic, Translational, and Applied Research. *Int. J. Plant Genomics* 2008:496957.

#### **Stakeholder Input.**

In addition to the outreach activity referred to above, we respond quickly, typically within 24 h, to direct inputs by email, phone or comments submitted from each page in the database. The Maize Genetics Executive Committee (MGEC) and our Working Group provide oversight and work with us on developing surveys of the community. In 2008-9, we asked for input from two groups, a Genome Browser Guidance Group, and many Beta testers.

#### **Acknowledgements.**

We are deeply grateful to our Working Group for their advice. Our current Working Group includes: Mihail Pop (Chief), Ed Buckler, Mike Freeling, Owen Hoekenga, Anne-Francoise Lamblin, Karen McGinnis, Lukas Mueller, Pat Schnable, Anne Sylvester, Doreen Ware and as *Ex Officio*, Volker Brendel and Marty Sachs. We met most recently April 2009, by teleconference and with all MaizeGDB staff convening at Ames, IA. For more details see [http://www.maizegdb.org/working\\_group.php](http://www.maizegdb.org/working_group.php).

For help in implementing the genome Browser, we greatly appreciate the Guidance Group: Peter Balint-Kurti, Sarah Hake, Damon Lisch, Mike Muszynski, and Virginia Walbot; and the Beta-testers: Alain Charcosset, Olivier Dugas, James Estill, David Hessel, Damon Lisch, Mike Muszynski, Paul Scott, Virginia Walbot, Rachel Wang and Cesar Alvarez-Mejia.

Submitted May 2009 by:

Mary Schaeffer  
Taner Z. Sen  
Ethyalinda Cannon  
Lisa Harper  
Carson Andorf  
Darwin Campbell  
Carolyn Lawrence (Lead Scientist)