MaizeGDB

Maize Genetics and Genomics Database

MaizeGDB STATUS REPORT

UPDATES, ACTIVITIES, AND NEW INITIATIVES

Prepared by: The MaizeGDB Team

Carson M. Andorf, Scott Birkett, Bremen L. Braun, Darwin A. Campbell, Ethalinda (Ethy) Cannon, Jack Gardiner, Lisa C. Harper, Carolyn J. Lawrence, Mary L. Schaeffer, Taner Z. Sen, and Jordan Thistle

DECEMBER 2010

Contact: C. Lawrence
USDA-ARS
1034 Crop Genome Informatics Laboratory
Iowa State University
Ames, IA 50011
Email: Carolyn.Lawrence@ars.usda.gov
URL: http://www.maizegdb.org
515-294-8280 (fax)
515-294-4294

TABLE OF CONTENTS

1 – Meeting agenda
Monday, December 6, 2010 (Times shown as Eastern time)

12:45 p.m. Dial in to the conference call or login to WebEx
    call: 1-800-857-3474   PASSCODE 33539
    WebEx: http://www.mymeetings.com/nc/join.php?sigKey=mymeetings&i=742984651&p=maizegdb&t=c

1:00 p.m.   Presentation: MaizeGDB's 2010 updates, accomplishments, and new activities

- Introductions
- Interface and tool development
- Data curation, outreach, and community support
- POPcorn
- Database and server management
- Upcoming
    - Team-building and reorganization
    - Interface redesign
    - Maize genome sequence stewardship: Looking to the future
- Charge to Working Group

2:00 p.m.   Working Group Executive Session
3:00 p.m.   Working Group Summarizes for the MaizeGDB Team
4:00 p.m.   Meeting adjourns

Working Group's Role
The Working Group is tasked with evaluating MaizeGDB current status and recommending a course of action that will insure that the MaizeGDB project tracks the trajectory of maize research as closely as possible. The ultimate goal of MaizeGDB is to provide a robust and timely source of data and analysis tools that will help researchers to investigate the biology of maize, both as a research model and as a crop.

Working Group Membership
Alice Barkan, Owen Hoekenga, Anne-Françoise Lamblin, Thomas Lübberstedt, Eric Lyons, Karen McGinnis, Lukas Mueller, Mihai Pop (chair), Marty Sachs, Pat Schnable, and Anne Sylvester.

2010 new members: Alice Barkan and Eric Lyons
Those rotating off the WG in 2010: Volker Brendel, Mike Freeling, and Doreen Ware

2 – Updates, activities, and new initiatives

Executive Summary
Since the Working Group Meeting in January of 2010, our focus continues to be on integrating new B73 assemblies and annotations into the MaizeGDB product. The MaizeGDB Genome Browser now has three views: BAC-based, B73 RefGen_v1, and B73 RefGen_v2 (available since May 2010). Links out to other databases' genomic views are in place, mechanisms for community members to contribute corrections to gene models are fully implemented (via a collaboration with V. Brendel's ZmGDB group), data including centromeres, markers/probes, gene models, and site-specific mutations are mapped, and additional data tracks and views are planned. A tool called uBLAST is being developed to help researchers prepare their large sequence-based datasets for inclusion in the MaizeGDB Genome Browser and is in the final stages of creating workflows for specific types of sequences. Views of gene models and associated data are now available. A nomenclature for gene model and transcript model names has been created by personnel at MaizeGDB and is pending Maize Nomenclature Committee approval in advance of implementation. As requested by the Maize Genome Sequencing Consortium, mechanisms to allow community members to document and gain access to assembly and annotation errors are in the planning stages. Surveys of the maize community indicate that functional annotation of the genome is the top priority, and the creation of a Maize Genome Annotation Consortium led by C. Lawrence with MaizeGDB as the focal point has been requested by the Maize Genetics Executive Committee. Training and outreach for researchers via on-site tutorials and videos accessible online are broadly attended and accessed. MaizeGDB's ancillary NSF-funded POPcorn resource now allows: 1. project search and browse capabilities, 2. sequence searches across many repositories, and 3. searches for data associated with sequence. Full integration of POPcorn within the MaizeGDB resource is planned for 2011. MaizeCyc, a comprehensive plant biochemical pathway database, is in the final stages of being implemented as the MaizeGDB Pathway View as a joint venture with P. Jaiswal from Gramene. A virtualized server environment is now fully implemented to ensure redundancy and availability of the MaizeGDB resource. With the outside funds currently available to MaizeGDB, our most significant plans in the near future are to remake the MaizeGDB interface based on recently developed technologies and user feedback over the course of the next two years. This plan ensures that the tools and data available to maize geneticists will remain up-to-date and relevant.

Personnel
Federal (* demarcates term appointments)
Carson Andorf, IT specialist (bioinformatics engineer), USDA-ARS in Ames, IA
      Lead programmer responsible for maintaining the MaizeGDB interface and the creation of data analysis tools unique to MaizeGDB. Also assists with the implementation of available software packages (like the MaizeGDB implementation of GBrowser for the Genome Browser and MaizeCyc for a Pathway View tool) and with server configuration and maintenance.
Bremen Braun*, IT specialist (programmer), USDA-ARS in Ames, IA
      Responsible for adding new data to the MaizeGDB Genome Browser and providing support, development and integration of new tools for MaizeGDB.
Darwin Campbell, IT specialist (database administrator), USDA-ARS in Ames, IA
      Responsible for the Oracle databases which are the backbone of the MaizeGDB project. Coordinates purchases (hardware, software, supplies/services) for the group, and is responsible for all associated accounting.
Lisa Harper*, geneticist (curator & outreach coordinator), half time, USDA-ARS in Albany, CA
      Responsible for community outreach and education, video tutorials, Phenotypic Controlled Vocabulary, Editorial Board management, and literature curation.
Carolyn Lawrence, research geneticist and lead scientist, USDA-ARS in Ames, IA
      Responsible for project management, coordination with outside groups, financial planning, and defining program direction in consultation with the MaizeGDB Team.

Mary Schaeffer, geneticist (curator), USDA-ARS in Columbia, MO
> Responsible for curation involving maps, loci, literature, QTL, etc. A member of the Maize Nomenclature Committee and a co-editor of the Maize Newsletter. Mary is an excellent source of historical information on why certain data are represented in a particular way.

Taner Sen, computational biologist, USDA-ARS in Ames, IA
> Responsible for the MaizeGDB Genome Browser, MaizeCyc, and other large-scale data management, analysis, and display endeavors.

State (funded by NSF)

Scott Birkett, POPcorn programmer, ISU in Ames, IA
> Responsible for the POPcorn sequence to publication search as well as the Palomero Toluqueño sequence public accessibility .

Ethalinda (Ethy) Cannon, POPcorn solution/application architect, ISU in Ames, IA
> Responsible for design and development of POPcorn website, its integration with MaizeGDB, and communication with groups providing data access.

Jack Gardiner, curator, ISU located on the University of Arizona campus, Tucson, AZ
> Curator on location at University of Arizona Campus, Tucson AZ: Responsible for identifying and recruiting gene expression, physical and genetic mapping, and epigenetic datasets with a special emphasis on large data sets.

Jordan Thistle, student hourly
> Created the RequestTracker instance that manages feedback and bugs, created the WordPress website for outreach, and is currently updating the Maize Genetics Conference abstract submission forms.

Davendra Jayasingam, student hourly
> Beginning some work toward the MaizeGDB interface redesign.

Noteworthy
- NSF has awarded funds for a proposal entitled Functional Structural Diversity Among Maize Haplotypes to Pat Schnable, PI. **C. Lawrence** (coPI) will get funds to add one additional computational curator to MaizeGDB for two years. That person's main focus will be on storing and displaying maize diversity data, especially copy number variations (CNVs) and presence/absence variations (PAVs). This additional focus on diversity creates an excellent opportunity to work collaboratively with CIMMYT as they develop their cyberinfrastructure to underlie survey sequencing for approximately 100,000 maize lines.
- Curators **Mary Schaeffer** and **Jack Gardiner** traveled to CIMMYT in June 2010 to discuss establishing areas of greater connectivity between CIMMYT and MaizeGDB that would allow better utilization of MaizeGDB by CIMMYT maize breeders. As an initial first step in creating stronger ties with CIMMYT, MaizeGDB agreed to develop curated linkages between MaizeGDB and CIMMYT for ~160 crop onology terms.
- In October 2010, Curator **Mary Schaeffer** attended the annual Biocuration meeting in Tokyo Japan and presented a poster and gave a talk on curation at MaizeGDB in general and on how MaizeGDB interacts with the maize community in particular. While in Japan, she also visited Oryzabase to gain insights on how they are handling phenotypic descriptors and attended a Rice Annotation Project policy session.
- **Carolyn Lawrence** was elected to the Maize Genetics Executive Committee in 2010
- We have published 3 peer-reviewed papers on MaizeGDB in 2010.

Response to January 2010 Working Group guidance
http://www.maizegdb.org/working_group.php
Lawrence

- **Visualization of BLAST hits that result in out-of-order alignments** to the chromosome is available via the MaizeGDB Genome Browser, though this requires visual inspection beyond the BLAST results page. We considered layering additional information to this effect onto the overview image carefully, but the range of possible BLAST results made a generalized solution to the problem impractical.
- **BLAST server performance has been improved.** The MaizeGDB BLAST service previously ran on a Dell E5504 server. This was a single processor machine with 4 cores @ 2.00 GHz and 8.0 GB of RAM. It ran a single apache web server. The current MaizeGDB BLAST service (which was made available in June of 2010) runs on a ProLiant DL785 G6 server. This machine has 8 CPUs with 6 cores per socket @ 2.792 GHz and 255 GB of RAM. It runs a mod proxy load balancer that distributes BLAST requests across 3 virtual machines each hosting an apache web server. The BLAST software has been upgraded from BLAST 2.2.17 to BLAST+ 2.2.23.
- **Complex phenotypic analyses via the current MaizeGDB interface are difficult.** For this reason, alternative mechanisms as well as collaborations to visualize, browse, and search phenotypes have been considered (see the Green et al. publication listed on p. 15) and phenotypic data representations will be a focus for the MaizeGDB Interface Redesign (see p. 13).
- **Phenotypic descriptions** need a unique accession identifier similar to a GenBank ID. At MaizeDB, the database ID's were shown. Note that these were not like GenBank ID's. Indeed, each phenotypic term does still have a database ID as it did at MaizeDB, but because there is no guarantee that the database ID's will be maintained over time we have discouraged their use as an identifier outside of the database itself. Like locus names, phenotypic identifiers are themselves unique. Endeavors to improve the Phenotypic Controlled Vocabulary itself are underway. A group of researchers to review that vocabulary as curator Lisa Harper tackles this challenge will be consulted regularly to insure that the final product is an appropriate representation of phenotypes measured by researchers. Use of the Phenotypic Controlled Vocabulary will be advertized widely once that project has reached a release stage, and a manuscript describing the vocabulary itself will be submitted to *SiGS* the data standards journal and official journal of the Genomic Standards Consortium.
- **The Maize Newsletter**'s Gene Review is now available from the front page of MaizeGDB. Because the Maize Newsletter is supported by an endowment to the University of Missouri, the suggestion for the MNL to be made available only online should be made to the editors (Jim Birchler and Mary Schaeffer).
- **Quantitative Trait Polymorphisms as well as diversity data** including presence/absence variations and copy number variations will be addressed in collaboration with the Schnable and Buckler groups. CIMMYT is planning a major endeavor in this direction to manage the informatics needs for survey sampling 100,000 lines, and we are in communication with Dave Marshall to insure that endeavors are coordinated. Mechanisms to view haplotypes within the MaizeGDB Genome Browser are planned. Because we are using GBrowse, various mechanisms for visualization already exist and can be deployed.
- **The ability to manage additional genome sequences for maize** exists, and using synteny browsers compatible with/developed for GBrowse we can allow comparative genomics views within the species. A nomenclature for gene models has been created by the MaizeGDB team that is expansible to lines other than B73 and groundwork is being laid for MaizeGDB to participate in managing improvements to the B73 sequence and its annotations that could be extended to other lines.

- Mechanisms will be put into place soon to document the **quality of the reference genome sequence** and a pipeline to get those improvements into GenBank have been suggested. For gene models, mechanisms to **improve the existing gene models have already been deployed** at MaizeGDB in collaboration with Volker Brendel's ZmGDB group and pipelines to improve the annotations based on reannotations submitted by community members are in the planning stages. In addition, as described in the MGEC report on the September 8, 2010 NSF meeting to discuss the concerns of and opportunities for the maize genetics community (available online at http://www.maizegdb.org/Sept2010MGEC@NSF.pdf), Lawrence is preparing a functional annotation RCN (Research Coordination Network) proposal to the Plant Genome Research Program.

The MaizeGDB Genome Browser
Sen, Andorf, and Braun

In continuing our efforts to develop the MaizeGDB Genome Browser, our main goal is to create a data-consistent view by displaying genomic data mapped to the most up-to-date B73 genome assemblies. In the past year, we have:

- Developed and released the MaizeGDB Genome Browser for B73 RefGen_v2.
- Obtained and integrated maize data aligned to B73 RefGen_v2 into database and created tracks in the MaizeGDB Genome Browser in collaboration with the MaizeGDB team to meet project plan objective 1B 36-month milestone. The data include B73 RefGen_v2 genome assembly from the Maize Genome Assembly Consortium; PUTs, EST, cDNA, GSS, microarray probes, Ac/Ds from PlantGDB; Uniform*Mu* from the Uniform*Mu* group; centromere and anti-CENH3 ChIP from the Jiang and Presting groups.
- Created a track showing the most recent gene models. Currently the gene structures available via MaizeGDB comprise the Release 5a.59 "Working Gene Set" product from the Maize Genome Sequencing Consortium, which is based on the B73 RefGen_v2 assembly.
- Developed interface improvements for the Genome Browser including the integration of existing pages with the Genome Browser for RefGen_v1 and RefGen_v2 (examples pages: locus, locus lookup, BAC, EST, cDNA, probes, MAGI redirects etc.).
- Updated the MaizeGDB BLAST tools functionality: integrated with RefGen_v2, added raw output option, upgraded to BLAST+.
- Serving two DAS (Distributed Annotation Server) tracks coming from PlantGDB: One track shows the quality of the Release 5a.59 gene models based on the evidence of cDNA and EST data. Another track shows the community annotation of genes.
- Created links to other databases that enable users to view the same regions of the maize genome on MaizeSequence.org, PlantGDB, and CoGE sites.

Future Plans:

- **Display genetic distance to physical distance:** Currently the browser notes how many base pairs are represented in a given screen. Soon, the approximate number of centiMorgans also will be noted based upon stored genetic maps.
- **GBrowse2** will be evaluated for a future implementation as the MaizeGDB Genome Browser.
- A track based on the **ISU IBM2009 markers** will be created in collaboration with the Schnable group.
- A **BIN track** will be created based on the ISU IBM2009 genetic map.
- A **Maize B73 Gene Expression Atlas Display** is to be deployed within the MaizeGDB Genome Browser in collaboration with the Kaeppler group.

Interface and tool development
Andorf, Braun, and Sen

- Created or modified 251 files related to the interface at MaizeGDB (these changes largely relate to improved or new functionality, interface changes, data changes, and bug fixes).
- Updated and tested interface to be compliant with PHP5 and Oracle 11g.
- Moved interface over to new virtual servers.
- Updated the "look and feel" for pages that are integrated with the MaizeGDB Genome Browser to handle views for multiple reference sequences.
- Updated the "look and feel" for the MaizeGDB general search. Twenty-five data sources are searched and retrieved in parallel. New search items include the following: synonyms, Genome Browser, Locus Lookup, and term definition.
- Consolidated the Locus Lookup tool to have a common interface for BAC-based and pseudomolecule assemblies.
- Implemented usage statistics on the reverse proxy to consolidate statistics for all MaizeGDB servers.
- Updated the MaizeGDB BLAST infrastructure: created two production BLAST servers and a backup server, virtualized the BLAST servers, created a mod proxy balancer (load balancing), created a BLAST data server that is mounted on each BLAST server, and created a backup BLAST database.
- Developed a reverse proxy server for the MaizeGDB infrastructure. This will permit the use of stable URLs that don't change when servers change.
- Implemented a memcache server to cache database queries to improve the speed of the web pages.

Future Plans:
- **uBLAST** - a tool to allow researchers to align their sequence on the B73 RefGen v2 assembly and create a track for the MaizeGDB Genome Browser. Once a track has been created and validated by the researcher, it can be submitted for inclusion in the MaizeGDB Genome Browser. This tool is in its final stages of development. The workflows are being created for mapping specific types of sequences.
- **MaizeCyc** - Developed by the project teams of Gramene and MaizeGDB databases in collaboration with the Maize Genome Sequencing Project, MaizeCyc is a catalog of known and/or predicted metabolic and transport pathways from maize. The 1.1beta version will be released in 2011. We are expecting the MaizeCyc database will undergo regular rounds of curation in order to integrate exciting new discoveries on biochemical pathways in maize. The computationally generated pathways will be subjected to manual curation based on published literature
- **Textpresso** - Textpresso is a tool that can be used by researchers as a search engine as well as by curators for literature-based data curation. Our curators will be able to go through the publications chosen automatically by keyword-based automatic queries to ensure the relevancy of publications prior to inclusion in the MaizeGDB resource then use the tool to pull out relevant data to be associated within the database. In addition, the tool will allow researchers to search available literature directly from the MaizeGDB resource.
- **See also upcoming items on p. 12.**

Data curation, outreach, and community support
Schaeffer, Harper, Gardiner, Campbell, Birkett, Andorf, Thistle, Cannon, and Sen

- Wrote definitions for 300 maize **phenotypes**, with expert help from researchers.
- Transitioned from using a **sequence update pipeline** from PlantGDB to in-house processing to allow updates directly from GenBank.
- Worked with CINVESTAV/Langebio to make the **Palomero Toluqueño sequence** available online. Transition of those data to GenBank is in process
- Collaborated with Truman State University, Kirksville, MO towards integrating their **GO annotation of gene models** into MaizeGDB. This annotation project (http://sam.truman.edu) is led by Diane Janick-Buckner and Jon Beck, and funded as part of an NSF shoot apical meristem microarray project (Mike Scanlon, PI). A copy of their MySQL annotation database, was provided for examination. Most of the annotations are generated over the summer via a summer intern program at Truman. The current plan is to review and accept annotation, with GO evidence codes for experimentally documented maize gene functions, and to share these annotations with the curated SwissProt/Uniprot group in Switzerland. This collaboration with Truman is expected to provide a model for accepting similar annotation from the community.
- Curated **1.67 millions** SNP provided by the Maize Diversity Project for HAPMAP1into the MaizeGDB Curation Database. Currently there are 27 lines, including B73. Data include genome coordinates on both v1 and v2 of the B73 reference sequence; the B73 context/flanking sequences; available genotypes for each of 27 lines; and the dbSNP submission ID, which serves as a link to dbSNP. While loading this very large data set was not an issue; efficient query capability is and options/solutions are currently being explored. Molecular breeders (at Missouri and CIMMYT) and other biologists have indicated that they minimally require files with the flanking sequences, to custom design probes and/or primers, and to be able to query on differences in genotypes between 2 inbred lines.
- Modeled and curated data from a Nimblegen-based **B73 gene expression atlas.** Shawn Kaeppler's group at the University of Wisconsin has surveyed gene expression for ~330,000 60-mer probes across 60 maize tissues. In collaboration with PLEXdb, MaizeGDB has been developing efficient and robust ways to associate this expression data with the current gene models to be displayed within the MaizeGDB Genome Browser and thereby add much needed expression information to the gene models. This "work in progress" includes appending plant ontology information to each of the 60 tissues that will be provided to the Plant Ontology Consortium when the data is released after the manuscript submitted by the Kaeppler group for publication is accepted.
- Collaborated with CIMMYT to include the **Crop Ontology** (CO) in MaizeGDB. This CO collaboration will feed into a larger ontology effort that supports the international germplam resources of the CGIAR (Consultative Group on International Agricultural Research). MaizeGDB has agreed to integrate their ontology information into MaizeGDB, link out to their associated data and include their data in MaizeGDB where appropriate, and to provide to them a list of traits in MaizeGDB that are not in the CO.
- Prepared **Maize News Letter 84** for redaction to hard copy and updated online posting (Jim Birchler and Mary Schaeffer, co-editors). The redaction is not funded by the USDA-ARS, but an endowment to the University of Missouri. There is some delay this year (owing to family health issues of the person charged with redaction).
- Updated the online MNL page at mnl.maizegdb.org/ to show **all available copies of MNL from 1932-2010**.
- Managed the 5-member **MaizeGDB Editorial Board**. Added data from over 60 publications recommended by that group.
- Developed a new WordPress-based **outreach site at MaizeGDB** and populated it with our video

tutorials, and links to other tutorials (outreach.maizegdb.org).

- Researched, designed, created, and distributed **4 tutorial/educational videos** about the maize genome sequence, assembly, annotation and use of the MaizeGDB Genome Browser, based on questions and input from scientists.
- Presented **live, on-site, formal tutorials** at University of California, Berkeley (10/2010), University of Arizona (4/2010), and Cold Spring Harbor Cereal Genomics Workshop (10/2009).
- Informal presentations at Stanford University, UC Berkeley, Iowa State University, and the USDA-ARS Plant Gene Expression Center.
- Answered "how to" questions by phone and email.
- Set up a request tracker to enable **consistent handling of user feedback, bug reports, and questions**.
- Administrated the **Maize Genetics Conference Steering Committee wiki**.
- Collected 307 abstracts and prepared the 219-page 2010 **Maize Genetics Conference program** and abstract book (Andorf and Schaeffer served as *ex officio* members of the 2010 Maize Genetics Conference Steering Committee). Developed a service where authors of posters or talks at the 2010 Maize Genetics Conference can upload their poster/talk to MaizeGDB to be displayed on reference pages.
- Administrated the Maize Genetics **Executive Committee mailing list**.
- Conducted Maize Genetics **Executive Committee annual elections**.
- Managed distribution of **email communications to cooperators** (13 this year).
- Created a **MaizeGDB Facebook page**, which has 82 "likes," 57 monthly active users, and 76 daily post views (as of November 30, 2010).

Database and server management
Campbell, Andorf, and Cannon

The MaizeGDB database contains nearly 60M records. 1.1M new records were added to the database during FY2010. The data transfer from PlantGDB to populate tracks for the MaizeGDB Genome Browser accounts for an additional 22.8M records (~82M total).

- Transitioned from physical servers to a **virtual server environment** for curation, development, and production. This entailed learning new systems architecture and planning the transitions with great precision to retain full stakeholder access to the database during the move.
- Worked closely with the Maize Genetics Cooperation – Stock Center personnel to standardize their code as we transitioned the Curation Databases from Oracle 9i to 11g. Their in-house developed code required updating to allow their interface to communicate with the Oracle 11g version of the database.
- Successfully **upgraded physical server hardware** while maintaining full web connectivity.
- Successfully installed Oracle Instant Client and associated software on the base configured virtual server, decreasing install time from many days to a few hours.
- Maintained **weekly database backups** to two off-site locations; one in Ames and one at the University of Missouri in Columbia, MO.
- Worked with Arturo Garcia to set up a **fully functional copy of the MaizeGDB Oracle Database at the University of Missouri in Columbia** for M. Schaeffer's use. The database is updated and managed remotely from Ames. This allows M. Schaeffer to test candidate data locally rather than over a network connection to the curation database in Ames, which facilitates a more better seamless transfer of data between Missouri and Ames.
- Integrated **Request Tracker server** and feedback into a web accessible management environment.
- Implemented a **WordPress Server for Tutorial/Outreach** accessibility.
- **Migrated group wikis to production virtual servers** from desktop PCs.
- **Migrated BLAST to a faster system** that allows load balancing.

POPcorn (http://popcorn.maizegdb.org**)**
Birkett and Cannon

Current status:
- **Collaborations and data-sharing** services with NCBI (Brian Smith-White), PlantGDB (Jon Duvick), BioExtract (Carol Lushbough), Dana Farber Cancer Institute (John Quackenbush), Grass Regulatory Information Server (Erich Grotewold), PLEXdb (Julie Dickerson), MaizeSequence.org (Doreen Ware), Photosynthetic Mutant Library (Alice Barkan), and Maize Assembled Genomic Islands (MAGIs) (Pat Schnable).
- **Release information**: Initial release occurred in April, 2010 and was announced to community in June, 2010
- POPcorn provides three search utilities:

    1. **Project Search** – search maize projects and resources. This was implemented by creating a set of data tables to hold information about ongoing and recent maize research projects and about active web resources that support maize research. These data are incorporated and updated through a set of curation tools created by the project
        o **Website**: http://popcorn.maizegdb.org/search/project_search/project_search.php
        o **Release information**: Initial release occurred in April, 2009. The database is updated monthly and currently holds information for 93 projects and 125 web resources.
    2. **Sequence-to-Sequence Search** – search for similar sequence across multiple data sets provided and hosted by multiple databases. There are currently 30 data sets hosted by 8 websites (DFCI, Grassius, MAGIs, MaizeGDB, NCBI, PLEXdb, PlantGDB, palomerotoluqueno.org). BLASTs are carried out via REST and SOAP Web Services.
        o **Website**: http://popcorn.maizegdb.org/search/sequence_search/home.php?a=BLAST_UI
        o **Release information**: Initial release occurred in April, 2010 and was announced to the maize community in June, 2010.
    3. **Sequence-to-Biology Search** – search for data linked to sequence. It is now possible to search for loci, variations, phenotypes, seed stock, genomic and transcript assemblies, transcription factors, and publications, as well as the capability to align input sequence on the MaizeGDB Genome Browser. Displaying BLAST hits on a whole genome view of the B73 assembly (see CViT, below) is under development.
        o **Website**: http://popcorn.maizegdb.org/search/sequence_search/home.php?a=SEQUENCE_SEARCH_UI

Future Plans:
- **CViT (Chromosome Viewing Tool)** – this open source tool was developed for the *Medicago* Sequencing Project and is being re-fitted to provide whole genome displays of aligned sequence, chromosome features such as centromeres and knobs, genetic and cytogenetic maps, and BLAST hits for the POPcorn and MaizeGDB sites. Because CViT reads the same data format as GBrowse, data that are presented as features on a chromosome or linkage group (e.g., images showing FISH to chromosomes) can be quickly prepared for display using a standard format (GFF) rather than requiring hand-drawn images. Data preparation can be done by MaizeGDB team members or by researchers themselves.
- Public releases of new website features (user interface improvements, new sequence-to-sequence and sequence-to-biology searches) happen every one to three months, roughly coinciding with the scheduled database update the first Tuesday of each month.

Upcoming: Team reorganization and interface redesign

Team reorganization: MaizeGDB functions as a team, though it is not necessarily clear to our stakeholders that this is the case. To allow researchers to contact appropriate individuals directly and to better convey the way we work, we propose to represent our team structure as shown in the following table. To improve our interactions, we also plan to begin weekly meetings of team leads in addition to our current weekly group conference call. A weekly journal club also will be created among the curators in an effort to increase awareness of what data are available and where the science is going.

Team Leads

Name	Role	Expertise	Contact
Carson Andorf	Bioinformatics engineer & programming lead	Interface design and development, tool development, web, network, and system administration	🌐 Contact_Icon
Darwin Campbell	Database administrator & operations lead	Data management, infrastructure hardware, facilities coordinator, purchasing	🌐 Contact_Icon
Carolyn Lawrence	Program direction & curation lead	creation/use of standards, evolution, map conversion (genetic, sequence, RN, cytological), community involvement	🌐 Contact_Icon
Taner Sen	Computational biology & 'omics lead	Genomic data integration, genome browser, gene annotation, proteins and protein interactions	🌐 Contact_Icon

Additional Personnel

Name	Role	Expertise	Contact
Scott Birkett	Programmer	NCBI Data Pipelining, POPcorn tool development and implementation	🌐 Contact_Icon
Bremen Braun	Programmer	Genome browser, tool development and implementation	🌐 Contact_Icon
Ethalinda (Ethy) Cannon	Programmer	POPcorn project planning, design, development, coordination with other data providers	🌐 Contact_Icon
Jack Gardiner	Curator	microarrays, gene expression, physical and genetic mapping, epigenetics	🌐 Contact_Icon
Lisa Harper	Curator & outreach coordinator	plant development, phenotypes, meiosis, cell biology, cytogenetics, outreach	🌐 Contact_Icon
Mary Schaeffer	Curator	biochemistry, genetic mapping, ontologies	🌐 Contact_Icon

Interface redesign: The MaizeGDB team is planning a complete interface redesign to keep the resource relevant and to allow for expansion in needed directions. The redesign will be both cosmetic and functional. The overall goal of the redesign is to create a clean modern interface with improved user interaction while maintaining the overall functionality of MaizeGDB. To insure that the new interface is useful, guidance and beta testing groups will be created and consulted (similar to how the MaizeGDB Genome Browser creation was managed). Listed below are concepts and technologies to achieve this goal:

- Redesign the look and feel: new colors, menus, and font consistency.
- Simplify page organization and navigation. Particular emphasis will be on phenotypic search and browse capabilities and incorporation of diversity data views.
- Improve the use of images, buttons and icons.
- Incorporate technologies to improve page load time: AJAX, CSS, caching, code compression.
- Improve the responsiveness of pages: AJAX, page flushing, caching.
- Remove or consolidate unused or outdated pages and/or data.

3 – Peer-reviewed publications authored by the MaizeGDB Team members since last meeting

Related to MaizeGDB

Harper, L.C., Schaeffer, M.L., Thistle, J., Gardiner, J., Andorf, C.M. Campbell, D.A., Cannon, E.K.S., Lawrence, C.J., and Sen, T.Z. The MaizeGDB Genome Browser Tutorial: One Example of Database Outreach to Biologists via Video. *Database* (Biocuration Virtual Issue) In preparation for December 2010 submission.

Schaeffer, M.L., Harper, L.C., Gardiner, J., Andorf, C.M., Campbell, D.A., Cannon, E.K.S., Sen, T.Z., and Lawrence, C.J. MaizeGDB: curation and outreach go hand-in-hand *Database* (Biocuration Virtual Issue) In preparation for December 2010 submission.

Green, J.M., Harnsomburana, J., Schaeffer, M., Lawrence, C.J., and Shyu, C.R. Multi-Source and Ontology-Based Retrieval Engine for Maize Mutant Phenotypes. *Database*. **Submitted:** 9/30/2010

Sen T.Z., Harper L.C., Schaeffer M.L., Seigfried, T.E., Andorf, C.M., and Lawrence C.J, Choosing a Genome Browser for a Model Organism Database (MOD): Surveying the Maize Community *Database*. 2010 10.1093/database/bap020.

Andorf C.M., Lawrence C.J., Harper L.C., Schaeffer M.L., Campbell D.A., and Sen T.Z. The Locus Lookup tool at MaizeGDB: identification of genomic regions in maize by integrating sequence information with physical and genetic maps. *Bioinformatics* 2010;26(3):434-6.

Other Publications by Team Members

Yennamalli R.M., Wolt, J.D., Rader, A.J., and Sen, T.Z. Thermophilicity in Endoglucanases is Fold-Specific. *BMC Structural Biology*. (**Submitted**: 6/8/2010)

International Arabidopsis Informatics Consortium. An international bioinformatics infrastructure to underpin the Arabidopsis community. *Plant Cell.* 2010;22(8):2530-6.

Lushbough C., Bergman M.K., Lawrence C.J., Jennewein D., and Brendel V. BioExtract server--an integrated workflow-enabling system to access and analyze heterogeneous, distributed biomolecular data. *IEEE/ACM Trans Comput Biol Bioinform*. 2010;7(1):12-24.

4 – Topics of requested input

1. What role should MaizeGDB play in supporting the call for functional annotation? As described in the MGEC report on the September 8, 2010 NSF meeting to discuss the concerns of and opportunities for the maize genetics community (available online at http://www.maizegdb.org/Sept2010MGEC@NSF.pdf), Lawrence is preparing a functional annotation RCN (Research Coordination Network) proposal to the Plant Genome Research Program.
2. Does the approach described in the presentation for updating maize sequence data errors and assemblies at GenBank seem reasonable?
3. Currently the MaizeGDB perspective on mapping RNAseq data to the genome for wide availability is to only allow atlas-level data. What is a reasonable way to address the community's need for storage of and access to RNAseq data?
4. Please comment on our new two-year plan to completely rework MaizeGDB's interface.