

## MaizeGDB Working Group Report 2006

Volker Brendel, Ed Buckler, Karen Cone, Mike Freeling, Owen Hoekenga, Lukas Mueller, Marty Sachs, Pat Schnable, Tom Slezak, Anne Sylvester, Doreen Ware

MaizeGDB is one of the most important resources for maize researchers, and it provides much of the glue that holds this international collection of scientists together. Excellent data has been well curated into MaizeGDB, and it is much more accessible than it was several years ago at MaizeDB. Overall, the MaizeGDB group has done a great job on limited resources. However, the next few years will provide major challenges to MaizeGDB, as the maize genome is sequenced and maize functional genomics research generates orders of magnitude larger datasets. MaizeGDB must play a central role in this revolution, to continue to serve the needs of the growing maize community. The working group identifies three central needs that will insure future MaizeGDB success:

### **(1) *MaizeGDB needs greater resources.***

This year 10.9 billion bushels of corn will be produced with a farm gate value of \$37.5 billion. Additionally, the value for ethanol production could double this value over the next decade. Currently, there are 3.5 employees devoted to the efforts of MaizeGDB, which is 0.0015% of the value of the corn crop (0.03% of the R&D effort). This is meager support for a database that is the central clearing house for genomic and genetic data for such an important crop. These resources will not allow the maize community to synthesize the results of genomics adequately and will impede the ultimate transfer of knowledge from the basic plant biologists in the lab out to the applied community and into the marketplace.

The working group noted that TAIR has 6 times as many employees as MaizeGDB, while the *Drosophila* database (flybase.org) has 9 times as many. Database improvement and maintenance are labor intensive and ongoing tasks requiring continued infusions of resources. Given the importance of corn production to the US economy as a whole, MaizeGDB needs a level of investment at least commensurate with *Drosophila* and *Arabidopsis* given the crucial mission to providing linkage of data for agronomic improvement.

Additional resources can come from at least three sources. First, federal interagency cooperation could provide increased resources. Second, the MaizeGDB group needs to take an active role in funding their research through competitive grants, as they are now beginning to do. Third, MaizeGDB needs to insure that any future federally-funded projects that propose to make use of MaizeGDB for lasting curation of valuable data must include budgetary support for any data submissions, curation activities that MaizeGDB is asked to provide.

At least doubling or tripling of funding through various sources would greatly enhance the effort by MaizeGDB at Ames, Iowa. An additional two programmers and three

curators at Ames would greatly accelerate the research. There are other bioinformatics groups around the country that can and are contributing to maize bioinformatics, but it is critical that MaizeGDB remain the vigorous leading group that synthesizes bioinformatics and makes it accessible to the entire research and agronomic community.

## **(2) *MaizeGDB needs to prioritize goals and future plans***

The MaizeGDB project has given careful thought to its future plans and asked for working group input on prioritization. The working group presents high and medium priority areas for consideration by MaizeGDB:

### *High Priority Areas*

- Community service should continue
  - Currently, MaizeGDB plays a central role in conducting central maize genetics community functions (ie. with annual meetings, votes, surveys, and the maize newsletter). This role should continue as it is critical to the success and cohesion of the research direction for the community. However, the working group recommends a careful self-evaluation of whether all actions are most efficient. For example, to optimize Trent Seigfried's time and make use of his skills, the group suggests that the Maize Genetics Meeting functions could be automated or otherwise not wholly dependent upon one of the two developers. It is important that MaizeGDB team continue to play this role, however, whenever possible they should get the curation resources from the community rather than providing them.
  
- Gene function prediction is important – Over the next few years, we will know the sequence of over 50,000 maize genes, but connecting genes to phenotypes is key to making genomics useful. MaizeGDB needs to play a leading role in curating, displaying, and analyzing the mutagenesis efforts in maize that will provide tools for functional analysis.
  - Curation of published analyses – Currently, MaizeGDB is the leading resource for published mutants of maize, which has nearly a 100 year history of research. These efforts take considerable curation effort and the working group encourage the MaizeGDB team to promote alternative approaches (eg., via outreach) and to develop automated approaches to mine literature (examples coming from the honey bee community).
  - Curation of stocks – Maize has a tremendous number and quality of genetic stocks. Continued curation of data regarding these stocks should be a top priority.
  - Several high throughput mutagenesis programs (both transposon and chemically-induced) are underway. It is definitely in the community interest that these studies be well curated at MaizeGDB. Many of these projects were funded through competitive funding, but funds were not requested for MaizeGDB curation. It may be necessary for these projects, MaizeGDB, and funding agencies to add resources for appropriate

- curation of these projects. Additionally, MaizeGDB needs to define the file formats for inputs, essentially reusable standards for data deposition. These standards should be readily accessible from the website.
- Integrated views and analysis of mutagenesis and phenotypic effects to gene are needed. These views need to be both from a genome view and from a pathway view. Competitive grants could provide resources to support these views, since the research questions are fundamental to basic functional genomics.
  - MaizeGDB needs to develop timelines for integration of the key datasets, and then publicize the timeline so that the research community knows when outcomes will be accessible.
- Maize Maps (Structural/Genetic/Cytogenetic) should continue to expand– MaizeGDB plays a key role in hosting and curating many of the maize genetic and cytogenetic maps that have been created over the last decade. However, as the sequenced genome for maize becomes available, the central maps will be changing to a B73 sequence. The main focus of MaizeGDB should be on linking relevant datasets to this sequence.
    - MaizeGDB should take a leading role in integrating the IBM genetic maps with the B73 genome.
    - When map data from the maize diversity study becomes available MaizeGDB should work with these researchers to develop a next generation genetic map.
    - Since there is substantial variation in maize genome structure, additional genomes will be sequenced and where possible coherent views need to be built in MaizeGDB.
    - MaizeGDB should focus on 2 major views - genetic and physical for now. Cytogenetic view although interesting and important has less general interest. Currently, there are often too many options, and users do not know what are the preferred maps.

#### *Medium Priority Areas*

- Gene structure prediction – As the genome is sequenced, many research groups around the nation are trying to predict gene structure and apply automated annotation of the genome. We do not believe that MaizeGDB should focus on doing this initial annotation but rather focus on:
  - Integrating, recording, and presenting the leading gene models (currently three). They should ensure their software can relate these gene models to the sequence centric genome views.
  - Organize experts in the communities to comment on the models (if funded as part of grant on this topic)
- Natural Diversity – Maize is the most diverse crop in the world, and USDA-ARS is responsible for the maintenance of this natural diversity. MaizeGDB needs to play a role presenting maize diversity and eventually helping plant breeders make use of this diversity. There are three other existing efforts ongoing in this area

(Panzea, GRIN, Gramene), and MaizeGDB needs to work with these groups to develop an efficient display of information. The working group suggests the following approaches to deal with intense the curation effort required.

- MaizeGDB should only curate QTL studies for germplasm that is held by the Maize Stock Center. Currently, the only mapping population with repeatable data held by the stock center is the IBM population.
- The IBM QTL studies should be curated with their phenotypic score data into the GDPDM schema that is currently used by the maize diversity group, wheat, rice, and sorghum groups. With these phenotypic scores, it will be possible to reanalyze the data as computational and genetic methodologies improve in the future.
- Several groups will begin mapping QTL down to the gene level over the next two years. These gene level QTL –trait associations should be integrated with the MaizeGDB gene function prediction work.

### ***(3) A Management Plan is needed.***

MaizeGDB is a young ARS group responsible for synthesizing, displaying and coordinating the outputs of the maize genetics/genomics community. Management of competing needs and setting clear objectives is the key to developing an ever improving MaizeGDB. It is essential that the MaizeGDB team set priorities to deal with the plethora of extant and emergent data. In particular, there needs to be coordination between the Iowa and Missouri teams in order to distribute the time efforts more efficiently to meet common goals. The working group suggests that the unified MaizeGDB team articulate together their goals in the short, mid and long term. Most importantly, the team needs to articulate how and when these priority goals will be met.

The working group recognizes that the priority goals suggested in this report require far more resources than are currently available. We suggest that two management plans could be devised: one that can be carried out within the budget limits of the current USDA-ARS commitment and a second that would be possible if appropriate resources were available. In this way, it will be transparent what is and is not possible within the resource constraints.

Short-term actions need to serve the long-term goals of MaizeGDB. The working group suggests that activities within the team need to be driven by a broader vision of the purpose and goals of MaizeGDB. For example, it seems that current action decisions are made based on urgency and the most persistent requests from the community. The team needs to take a leadership role in defining how, when and which datasets should be curated. If there are commonly agreed short and long-term actions, then all action decisions can be made based on helping to move MaizeGDB towards the long-term goals.

In summary, the working group makes the following suggestions for developing a management plan:

- Datasets to be curated and displayed need to be prioritized. Participant actions and timelines must be established in a systematic fashion.
- Numerous groups would like to have their data displayed at MaizeGDB, but MaizeGDB should feel free to say no if these data do not match the long-term goals.
- As in any scientific endeavor, it is important to devote substantial time to development of the newest views or approaches rather than just putting out fires.
- The PoPcorn portal was very interesting, and we encourage MaizeGDB to develop an independent and diverse portfolio of competitive grant ideas and funded projects.
- The management plan should be in written form and consistent with realistic expectations.