The MaizeGDB Genome Browser Survey Results

Taner Z. Sen[1], Trent E. Seigfried[1], Darwin A. Campbell[1], Mary L. Schaeffer[2], Lisa C. Harper[3], Carolyn J. Lawrence[1]

[1]USDA-ARS, Ames, IA [2]USDA-ARS, Columbia MO [3]USDA-ARS Albany, CA

Abstract

MaizeGDB (http://www.maizegdb.org) is the community database for maize genetics and genomics. Based upon the 2006 MaizeGDB Working Group Report and the Allerton Report, it has become evident that the focus of MaizeGDB must be shifted to better accommodate a sequence-centric paradigm. In order to (1) show how the data at MaizeGDB relate to the maize genome, (2) relate MaizeGDB's current sequence data with other sequence information as it becomes available, (3) become the keeper of maize's 'official' set of gene models (which will enforce proper nomenclature), and (4) create a way to compare the various assemblies and annotations simultaneously, the feasibility of implementing of a genome browser at MaizeGDB was investigated and various available genome browser software were evaluated. Because the maize community communicates well, has a clear vision of their research problems, and has good ideas on how best visualize a sequenced maize genome, we prepared a survey to gauge cooperators' impressions of existing software and to find out what sorts of functionalities they would like to have in the MaizeGDB Genome Browser. Here we present the results of that survey as well as the rationale for why we chose GBrowse as MaizeGDB's Genome Browser software, and the roadmap for implementation.

The Need for a Genome Browser

Per the Working Group's Guidance:

- "MaizeGDB needs to play a leading role in curating, displaying, and analyzing the mutagenesis efforts in maize that will provide tools for functional analysis."

- "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."

- "…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."

- "As the genome is sequenced, many research groups around the nation are trying to predict gene structure and apply automated annotation of the genome. …MaizeGDB should 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."

Per the Allerton Report:

- "MaizeGDB should become the centralized sequence resource soon after the genome is complete (2009–2010)."

- "To improve access to maize sequence data, resources that integrate various gene models and annotation sets must be made available to MaizeGDB."

- "The sequenced maize B73 genome holds great promise for contributing to basic and translational research. To take full advantage of that promise we need to 1) make MaizeGDB the centralized sequence resource, 2) make plans and implement next level map-based sequencing efforts, 3) provide increased capacity and lower costs for maize transformation technology, 4) increase sequence-based reverse genetics, 5) coordinate expression platforms so all data are easily shared, 6) conduct a major phenomics effort that is effectively integrated, and 7) increase participation in maize research."

- "Convert MaizeGDB to a sequence-oriented database – 2011 (3 years)"

Preparation of the Survey

Preparation of the Survey

- The MaizeGDB team created and distributed the survey on behalf of the Maize Genetics Executive Committee.

- After we prepared the survey, we sent it to the Working Group and Executive Committee for suggestions and incorporated the suggestions into the survey.

- We appreciate that the Executive Committee supported the survey and promoted it among maize cooperators.

Eliminating Bias from the Survey

Recommendations from Assist. Prof. Patrick Armstrong from ISU-Psychology:

1. Instead of asking people to circle a specific website that uses a specific browser, put the websites in alphabetical order, so that people will be able to find the websites they are using very easily without thinking which Genome Browser the website is using.

2. Combine ranking of features with rating (basically assigning weights to specify how important a specific feature is to the user)

Choosing the survey takers

- We sent the survey to 1,241 Maize Cooperators

- Definition of a cooperator:
  - Attendees of maize meetings
  - Researchers publishing frequently on maize
  - People who specifically requested to be considered a maize cooperator from MaizeGDB

- Privacy: we used a randomly generated key that is deleted at the end of the survey.

- Uniqueness: key system ensured that each cooperator can take the survey only once.

Survey Results

Time Spent Accessing Maize Online

- 37%: 1 or 2 hours.

- 39%: 2- 5 hours.

- 15%: >5 hours.

- Only 8% of the survey takers did not use online maize data resources.

Genome Browsers used

- 66% of the respondents use Maizesequence.org and the number is the same for Gramene users. A total of 75% use either MaizeSequence or Gramene.

- Although both sites use Ensembl as a genome browser, only 26% of the respondents acknowledged they are using Ensembl. This result shows that the users may not be aware of the underlying software for browsers that the various websites are using.

Genome Browsers Used

- TAIR, MAGI, and PlantGDB are being used by 54% of the respondents (but not always by the same people).

- 42% use NCBI's Map Viewer.

- As above, although 45% use TAIR, only 22% report that it is using GBrowse.

Feature Rankings

- The features are sorted as follows (rankings are shown in parentheses where a lower number indicates more support):
  - Ease of use (1.9)
  - Visuals (2.6)
  - Speed (3.2)
  - Cross-species comparison (3.7)
  - Multiple gene selection (4.1)
  - Differentiation between comput. and exp. data (4.1)
  - Ontologies (5.1)

- Clearly, the respondents want an intuitive genome browser that allows researchers to locate the needed data in the most accessible and fastest fashion.
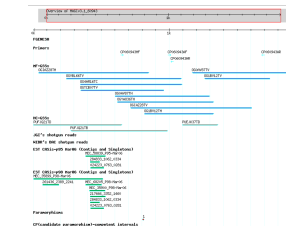
Desired Features

- The most desired feature section of the survey is very helpful to guide genome browser developers in the creation of new features. The users want
  - Ability to reach specific data using the most intuitive tools.
  - Downloadable data sets in various formats.
  - Enhanced cross-referencing between different websites.
  - The most current data and the tools that are easy to learn and to apply for their specific research needs.

- In short, the users want the minimized hassle and effort in reaching the needed maize data in fastest fashion.

Bad Genome Browser Examples

- Among 29 comments left in "Bad genome browser examples", 19 of them cite either Maizesequence.org or Gramene (66%), which use Ensembl as their genome browser.
- The reason might be that Maizesequence.org or Gramene is the most used browser for the maize cooperators (75% of the respondents uses either site), but the high percentage of those discontent hints that some issues with Ensembl may need to be addressed.
- The respondents usually cite
  - the perceived slowness of the website as the major problem
  - "many, many non-intuitive steps to get information"
  - "too many choices, there should be a sensible default for newcomers"

Conclusions and Roadmap

Conclusions

- Based upon our analysis, choosing **GBrowse** will provide maize researchers the following strengths:
  - Ease of use, visuals, speed
  - Community development
  - Comparative genomics tools

- Providing interlinks between MaizeGDB and Gramene will enrich the research tools with minimal cost.

A **snapshot from the MAGI website showing MAGIv3.1_60943 region using GBrowse**. Eddy Yeh and Karthik Viswanatha from the Schnable Group will assist MaizeGDB during the genome browser implementation when we encounter roadblocks. The MaizeGDB Genome Browser will also provide various maize assemblies side-by-side.

Roadmap for the Genome Browser

1. We plan to start implementing the Genome Browser right away.

2. We plan to get a copy of the database from maizesequence.org so that we can start working on how the sequence data will fit within the database schema of Gbrowse and to get available gene models from maizesequence.org, Dana Farber, Schnable, and Brendel.

3. We plan to choose 5 people to provide guidance (from academia, government, and industry with at least one representative from outside of the US), and 10 people for beta testing among the cooperators.