Database, Vol. 2010, Article ID baq007, doi:10.1093/database/baq007

Original article

Choosing a genome browser for a Model Organism Database: surveying the Maize community

Taner Z. Sen[1,2,*], Lisa C. Harper[3,4], Mary L. Schaeffer[5,6], Carson M. Andorf[1], Trent E. Seigfried[1], Darwin A. Campbell[1] and Carolyn J. Lawrence[1,2]

[1]USDA-ARS Corn Insects and Crop Genetics Research Unit, [2]Department of Genetics, Development and Cell Biology, Bioinformatics and Computational Biology Program, Iowa State University, Ames, IA 50011, [3]USDA-ARS Plant Gene Expression Center, 800 Buchanan Street, Albany, CA 94710, [4]Department of Molecular and Biology, University of California Berkeley, Berkeley, CA 94720, [5]USDA-ARS Plant Genetics Research Unit and [6]Division of Plant Sciences, University of Missouri, Columbia, MO 65211, USA

***Corresponding author:** Tel: +1 515 294 5326; Fax: +1 515 294 8280; Email: taner.sen@ars.usda.gov

Submitted 16 November 2009; Revised 8 March 2010; Accepted 9 March 2010

As the B73 maize genome sequencing project neared completion, MaizeGDB began to integrate a graphical genome browser with its existing web interface and database. To ensure that maize researchers would optimally benefit from the potential addition of a genome browser to the existing MaizeGDB resource, personnel at MaizeGDB surveyed researchers' needs. Collected data indicate that existing genome browsers for maize were inadequate and suggest implementation of a browser with quick interface and intuitive tools would meet most researchers' needs. Here, we document the survey's outcomes, review functionalities of available genome browser software platforms and offer our rationale for choosing the GBrowse software suite for MaizeGDB. Because the genome as represented within the MaizeGDB Genome Browser is tied to detailed phenotypic data, molecular marker information, available stocks, etc., the MaizeGDB Genome Browser represents a novel mechanism by which the researchers can leverage maize sequence information toward crop improvement directly.

Database URL: http://gbrowse.maizegdb.org/

Introduction

A genome browser is to genomic sequence data as a web browser is to the World Wide Web: both offer logical access to datastreams that are otherwise unintelligible. With the advent of new DNA sequencing technologies and the availability of copious amounts of sequence-based data from many species, genome browsers have been developed as a means for researchers to view, interact with, search through and display sequenced genomes as well as to compare syntenic or similar regions of genomes among related species. Various genome browsers have been created over the years, each with particular strengths and weaknesses. Many provide independent solutions for integrating and visualizing sequence-based data alongside genetic and phenotypic information.

Community resources including Model Organism Databases (MODs) [e.g. TAIR (1), FlyBase (2), etc.], Clade-Oriented Databases (CODs) [e.g. Gramene (3), SGN (4), etc.], Automatic Annotation Shops [e.g. PlantGDB (5), JCVI (6, 7), etc.] and others have a responsibility to provide timely access to sequence data well-integrated with existing traditional biological data. Determining how best to choose genome browser software to meet the needs of users within the context of a group's maintenance capabilities is a major challenge for the groups working to build and maintain these community resources. Described here are the methodologies we used to determine which

Database, Vol. 2010, Article ID baq007, doi:10.1093/database/baq007

genome browser to implement at MaizeGDB (8–10), the MOD for maize.

The need for a genome browser at MaizeGDB

These are exciting times for maize researchers and breeders. Not only is maize a major crop worldwide; a reference genome sequence for the inbred line, B73, has been released [www.maizesequence.org; (11)]. As of August 2009, the minimum tiling path included 16 910 sequenced Bacterial Artificial Chromosome (BAC) and fosmid clones and encompassed 2.12 Gb or 93% of the 2.3 Gb B73 genome (12). The B73 pseudomolecules (12) are available through the Arizona Genomics Institute website (http://www2.genome.arizona.edu/genomes/maize). Other whole-genome sequences include the shotgun sequences of an ancient popcorn landrace, Palomero Toluqueño (13) and the maize inbred line Mo17 (from JGI- the Joint Genome Institute, with D. Rohksar leading the group, http://www.phytozome.net/). In addition, an extensive haplotype map has been published for 27 lines of maize, enabling researchers to establish novel relations between genetic, physical and diversity data (14, 15). Other sequence-based resources include over 2 million public ESTs (http://www.ncbi.nlm.nih.gov/dbEST/dbEST_summary.html) and a large number of genic sequences from gene-enriched libraries (16, 17). Various research groups and consortia integrate large portions of these data sets, each in their own way. Examples include PlantGDB [(5); www.plantgdb.org], the Dana Farber [http://compbio.dfci.harvard.edu/tgi/tgipage.html; (18)], MAGI [http://magi.plantgenomics.iastate.edu/; (19)], NCBI RefSeq (20) and Uniprot (www.unitprot.org; The UniProt Consortium 2009). Integration of the large data sets, at a single location, with the information about the position, orientation and sequence of genes, genetic markers, variations and their association with phenotypic data would allow for a detailed understanding of the maize genome within its biological context, when presented as centrally accessible and simultaneously viewable.

At the completion of the Maize Sequencing Project, it is anticipated that genomic data and gene models will be transferred from the Maize Genome Sequencing Consortium's project database MaizeSequence.org to MaizeGDB (8–10) and Gramene (3). As a federally funded, long-lived resource, MaizeGDB is tasked to serve maize geneticists' and breeders' longitudinal data access and analysis needs. To accomplish these tasks, MaizeGDB primarily relies on direct participation by members of the maize research community including the Maize Genetics Executive Committee (MGEC; a group tasked to identify both the needs and the opportunities for maize genetics and to communicate this information to the broadest possible life science community), the MaizeGDB Working Group (a panel that offers guidance for MaizeGDB's continued development), and direct interaction with individual researchers. Other databases, such as TAIR (1) and SGN (4) also rely on similar means to interact with and receive feedback from their communities. However, to the best of our knowledge, the MaizeGDB Working Group is fairly unique for a few reasons: the group (i) meets at least once yearly: many other database groups' advisory boards are formed then fail to meet, (ii) documents guidance online (see http://www.maizegdb.org/working_group.php) and (iii) routinely allows representatives from other database groups and various funding agencies to observe their meetings. The successful guidance provided by the MaizeGDB Working Group has even inspired others including Soybase (21) and GRIN (http://www.ars-grin.gov/npgs/) to create similar guidance committees.

Currently, MaizeGDB stores information on: loci (genes and other genetically-defined genomic regions including QTLs), variations (alleles and other sorts of polymorphisms), stocks, molecular markers and probes, sequences, gene product information, phenotypic images and descriptions, metabolic pathway information, reference data and contact information for maize researchers. Like many other MODs [such as TAIR (1), Oryzabase (22) and Soybase (21)], MaizeGDB incorporates and integrates newly generated genomic data into its existing database and develops tools to help visualize genome structure, gene models, functional data, and genetic variability. Toward this end, two groups directed MaizeGDB to evolve a more sequence-centric paradigm: the MaizeGDB Working Group (via their 2006 guidance document; see http://www.maizegdb.org/working_group.php) and maize principal investigators (in the 2007 Allerton Report that documents outcomes of a special 2-day gathering of maize community with a focus on 'The Future of Maize Genetics Planning for the Sequenced Genome Era'; see http://www.maizegdb.org/AllertonReport.doc). The time was right to carefully consider implementing a genome browser as a way to integrate genomic sequence features with the existing genetic and physical information at MaizeGDB.

When we began considering the implementation of a genome browser at MaizeGDB, various other resources already represented maize genomic sequence visually via genome browsers. Most notably, MaizeSequence.org, Gramene (which removed their maize-centric genome browser when MaizeSequence.org was released), PlantGDB with its maize ZmGDB browser and the Maize Assembled Gene Islands (MAGI) resource. A specific challenge for MaizeGDB was whether to follow the lead of the Maize Genome Sequencing Consortium and collaborate with that group to further develop MaizeSequence.org. This would be an efficient use of funds in the short term given that both groups could collaborate to maintain a single maize genome browser.

Another issue for consideration was the MaizeGDB team's charge to make decisions based upon input from the maize community. We are fortunate at MaizeGDB to serve a remarkably cooperative community that communicates well. This time honored tradition of communication and cooperation goes back to 1929 when R.A. Emerson held an informal 'cornfab' gathering in his hotel room with maize researchers during the American Association for the Advancement of Science (AAAS) meeting (23). This meeting led to the creation of the Maize Genetics Cooperation (MGC) as well as the publication of the MGC Newsletter (24) and formed the foundations of the MGC – Stock Center, which is one of the finest examples of cooperative resource sharing (23). The MaizeGDB team continues in this tradition by facilitating online mechanisms for continued communication.

We followed the hierarchical strategy below to gather the information needed to determine how to proceed with potentially implementing a MaizeGDB Genome Browser:

(1) Should MaizeGDB make a genome browser available at all? If researchers were happy with the existing options, implementing another resource would be a waste of time and resources.
(2) If researchers wanted MaizeGDB to implement a genome browser, we needed to know:
    (a) what they liked and did not like about available maize genome browsers and
    (b) examples of workflows they would like to be able to carry out so that we could evaluate which software could best meet our stakeholders' needs.

With these ideas in mind, we approached the MGEC and MaizeGDB Working Group. These groups offered to survey the maize community on our behalf and worked with us to prepare a survey that aims to answer questions 1, 2a and 2b.

Materials and methods

Preparation of the survey

The MaizeGDB team prepared an initial draft survey, and sent it to the MaizeGDB Working Group and the MGEC for suggestions. The updated set of questions was considered by Dr Patrick Armstrong in the Department of Psychology at Iowa State University who made recommendations on how to eliminate potential bias by the wording and ordering of questions. The final form of the survey can be found at http://www.maizegdb.org/browser_survey/ and in the Supplementary Materials section of this document.

In November 2007, MaizeGDB personnel distributed via email a request by the MGEC for all 'maize cooperators' (totaling 1241 at that time) to take a survey regarding their use of online maize data resources with emphasis on browsing the available genome sequence. 'Maize cooperators' are a list of maize researchers maintained at MaizeGDB and include attendees of maize meetings, researchers publishing frequently on maize, and any persons who specifically request to be considered a maize cooperator. Each cooperator received a randomly generated unique key to ensure that each email recipient was only able to submit answers to the survey once.

The number of respondents. Among the 1241 cooperators surveyed, 99 responded. This number is comparable to the number of participants to the last MGEC membership election where 234 of the 1190 contacted cast a ballot. Because the Genome Browser Survey requested detailed answers to the researchers' needs and not every maize researcher would feel knowledgeable on genome browsers, this level of response to the survey exceeded our expectation.

Results

The raw survey results can be found in the Supplementary Material section, as well as at http://www.maizegdb.org/browser_survey/analyze.php. Tabulated results are located at http://www.maizegdb.org/browser_survey/analyze-tab-delimited.php.

Time spent accessing maize data online

Thirty-seven percent of the survey takers reported that they spend an hour or two each week online to access maize data. Thirty-nine percent spend between 2 and 5 h. Fifteen percent spend >5 h online to access maize data. Only 8% of the survey takers did not use online maize data resources.

Genome browsers used

Sixty-eight percent of the respondents reported that they use MaizeSequence.org and 66% use Gramene. A total of 76% use either MaizeSequence or Gramene. Although both sites use Ensembl [one genome browser software option; described in the 'Discussion' section; (25)] as their genome browser, among the users of these websites, only a total of 35% of the all respondents acknowledged using Ensembl. This result shows that users may not be aware of the underlying browser software that the various websites use.

MAGI and PlantGDB are being used by 54% of the respondents (but not always by the same people). A total of 42% use NCBI's Map Viewer. As above, although 45% use TAIR, among these users, only 31% acknowledged that they are using GBrowse [another genome browser software option; described in the 'Discussion' section; (26)].

Database, Vol. 2010, Article ID baq007, doi:10.1093/database/baq007

Table 1. Main genome browsers listed in alphabetical order, their focus and the example databases that use them

Genome browsers	Focus	Link	Example databases
Ensembl	Comparative Genomics, mainly for CODs, but also MODs	http://www.ensembl.org/	Gramene
GBrowse	MODs with some comparative genomics	http://gmod.org/wiki/Gbrowse	TAIR, Flybase, SGN
NCBI Map Viewer	Browsing all biological sequences stored in the NCBI	http://www.ncbi.nlm.nih.gov/projects/mapview/	NCBI
UCSC	Vertebrates and non-vertebrates, both for MODS and CODs	http://genome.ucsc.edu/	Human Genome Project
xGDB	Customized to work with different types of data	http://xgdb.sourceforge.net/	PlantGDB

Note that the NCBI Map Viewer is not available to be downloaded on local machines.

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 computational and experimental data (4.1) and ontologies (5.1). Clearly, the respondents want a genome browser that allows them to find data quickly and easily.

Desired features

The 'desired features' section of the survey should be very helpful to guide genome browser developers in the creation of new features. Survey respondents expressed interest to reach specific data using the most intuitive tools that require short learning time. They also reported a need for enhanced cross-referencing between different websites and called for downloadable data sets in various formats. In short, respondents want minimized hassle and effort in reaching needed maize data.

'Bad' genome browser examples

We asked respondents about what they do not like about current genome browser examples to give us an indication of browsers or options to avoid. 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 the discontent hints that real issues may lie with some features of Ensembl that need to be addressed by its developers. The respondents usually cited the perceived slowness of the website as the major (and sometimes the only) problem. Another reported problem was, to quote one respondent, 'many, many noninituitive steps to get information'.

Four software suites to choose from

Although many genome browser software platforms exist, survey respondents were most familiar with Ensembl, GBrowse, the NCBI Map Viewer (27), the UCSC Genome Browser (28) and eXtensible Genome Data Broker (xGDB) (29). Each genome browser is designed with a different focus (Table 1). Here, we provide a short review of some of their functionalities we considered in choosing a genome browser for MaizeGDB. Among these genome browsers, the NCBI Map Viewer is not downloadable to local machines, so it was not considered as a choice for the MaizeGDB Genome Browser. Because our users are extensively using the NCBI Map Viewer, we include it in our review for comparison.

Ensembl

The Ensembl browser was originally developed to manage and display genomic data for the Ensembl project as a human genome browser (25). Initially, the developers focused on mammalian genomes, but now Ensembl include plant genomes. Some examples include the plant Ensembl genomes portal (http://plants.ensembl.org/), Gramene and atEnsembl. Ensembl especially excels in comparative genomics visualization and analysis. It provides a flexible framework that displays a wide variety of genomes [currently the Ensembl browser displays 48 genomes (25)]. A recent addition to Ensembl is the new multiple alignment pipeline that passes data through three different programs [the Enredo–Pecan–Ortheus (EPO) pipeline (25, 30, 31)] to obtain alignment results.

Ensembl's web interface combines many distinct, dynamically-generated views (e.g. genes, maps, contigs) to address different needs of the researchers. The framework is also integrated with multiple tools, including the similarity search tools BLAST and SSAHA, the retrieval software EnsMart and the Distributed Annotation System (DAS) framework (32, 33) for sharing and displaying distributed data sets on any publicly available Ensembl instance

(i.e. locally installed software). Ensembl is designed to be portable—users with advanced programming skills can extend or modify Ensembl code through the Ensembl API (application programming interface), a downloadable open-source package.

GBrowse

The Generic Model Organism Database Project (GMOD) (http://gmod.org) has the mission to build tools designed to serve the needs of MODs. One of the major and most popular tools developed by GMOD is the Generic Genome Browser (GBrowse) (26), an open-source web-based framework for displaying genomic annotations and features. Similar to other genome browsers, GBrowse allows the user to scroll and zoom within a genomic region, search for features based on name or keyword search and customize feature tracks. A useful visual element in GBrowse is that each feature type can be represented by various customizable 'glyphs', which are essentially symbols that vary in shape, color and size to represent genomic elements.

GBrowse was designed to be portable and extensible (i.e. its code is modifiable to add new capabilities). A developer can modify GBrowse at the following three different layers: the database layer, the data model layer and the application layer. This flexibility allows the administrator to control how the data are stored, how the data are visualized, and how the user interacts with the data. GBrowse is a downloadable, stand-alone, open source package and was designed to facilitate third-party plug-ins for data analysis and visualization. Some examples include plug-ins for calculating linkage disequilibrium, dumping data as GFF or FASTA and facilitating the connection between GBrowse and Galaxy (34). GBrowse can also be integrated with the comparative map viewer CMAP (35), the BioMart data mining system (36) and the TextPresso text mining tool (37). Some developers have even harnessed the GBrowse extensibility to create a web server for GBrowse that allows access without the hassle of local installation (38). Similar to the Ensemble browser, users can upload custom data (flat files or an URL) with ease through the DAS platform (32, 33), which decentralizes data storage by allowing the display of third-party annotations. GBrowse is used by many MODs, including TAIR (1), WormBase (39), and Mouse Genome Informatics (MGI) (40), as well as CODs, such as SOL Genomics Network (4). The International HapMap project (http://hapmap.ncbi.nlm.nih.gov/) also uses GBrowse as their genome browser.

NCBI map viewer

As a static repository, the National Center for Biotechnology Information (NCBI) strives to preserve the archives of large species-specific data sets for the scientific community. Its primary mission is to keep them up-to-date, searchable, and publicly available. NCBI accomplishes these herculean tasks in collaboration with many researchers and curators across species. NCBI also provides a range of tools for the visualization and analysis of genomes. Central to these tools is its genome browser, Map Viewer (27). Map Viewer is not designed to be customizable, but it is capable of visually representing maps and genomic elements and providing links to the web pages that include the most current and comprehensive data about these genomic elements.

One of the main disadvantages of Map Viewer is that it does not have the capability to be downloaded and installed to personal servers. It is specifically designed to work under the NCBI framework.

The UCSC genome browser database

The University of California Santa Cruz (UCSC) Genome Browser Database (28) started as part of the Human Genome Project (41) to make newly generated human genomic sequences publicly available. Although the UCSC genome browser remained focused on the human genome, its content over the years has extended to a cross-comparison platform of 19 vertebrate and 21 invertebrate species (28). The UCSC browser currently serves many tracks including an evolutionary conservation track based on 28 species, variation and disease tracks, and mammalian gene collection tracks. Although plant genomes are not included on the main UCSC site, any genome sequence can be uploaded to a locally installed instance of the UCSC browser. An example is the Joint Genome Institute's (JGI) 'tree of life' (http://genome.jgi-psf.org/). The browser's code is open-source; therefore, customization by developers is possible. Also, the browser allows 'custom tracks' that may be uploaded to the UCSC website or to any available instance of the UCSC genome browser using the DAS framework. Similar to GBrowse and Ensemble, DAS tracks in the UCSC browser can be created temporarily on any instance that uses the DAS framework, but these tracks will only be privately available for the user who uploads them. It is also possible to use the DAS framework to create publicly available, permanent tracks to display data provided by third-party servers, but this requires access and administrative privileges to the main server where the genome browser is located.

eXtensible Genome Data Broker

The xGDB (29) is the genome browser developed by personnel working at PlantGDB (5) to facilitate their need for a system to manage, store and display genomic evidence for 16 green plant genomes (including the maize genome). xGDB is a software package designed to view the outcomes of sequence analyses within a genomic context. xGDB can be customized for various individual research tasks and analysis needs. Other features of xGDB include search tools, online publishing, web services and third-party tool

Database, Vol. 2010, Article ID baq007, doi:10.1093/database/baq007

integration. The browser serves data through the DAS framework.

Technical requirements for implementing a genome browser

The basic technical requirements for implementing each of the browsers are very similar. A basic understanding of the operating system (e.g. Linux, Mac OS or Windows) and how to execute basic command line commands is helpful. Each browser has step-by-step documentation on how to install the software, but occasional troubleshooting is required. This generally requires installing additional software packages and resolving dependency issues. Most of the browsers either require or recommend setting up a back-end database. Basic knowledge of how to create, populate and maintain a database may be required. The MySQL database is the most common database used by the browsers, but there is limited support for Oracle, Chado, PostgreSQL and other databases.

Download/upload data capabilities are very similar across browsers. The data for display in the genome browsers accepted in GFF (General Feature Format) across the board, with support for other data formats as well: WIG and SCF for GBrowse; GTF, PSL, BED, BedGraph, WIG for Ensembl. The UCSC Genome Browser is the most flexible, accepting GTF, PSL, BED, BedGraph, WIG, as well as bigwig, MAF and microarray data formats. Meanwhile, xGDB only accepts GFF and XML formats. However, it should be noted that an experienced programmer can easily write an 'adapter' for any genome browser to accept customized or idiosyncratic data formats.

The programming skills needed to setup and maintain a genome browser are minimal. This involves setting up modules (like Perl) and executing scripts. Intermediate programming knowledge may be necessary to import data into the database. The skill level is dependent on the complexity of the data. Most browsers provide scripts for commonly formatted data (e.g. GFF). Customization of the browser (colors, fonts, sizes, etc.), requires knowledge on how to update simple HTML (Hyper Text Markup Language) and CSS (Cascading Style Sheets) code. For more advanced customization, a more in-depth understanding of web technologies may be needed (HTML, Perl, PHP, CSS, JavaScript, etc.). For all the genome browsers reviewed here, it is also easy to create links to internal web pages. This is especially helpful for MODs who aim to integrate a genome browser with existing data displays.

The 'Next-generation' browsers

Though not part of our survey, it is worth mentioning some new 'next generation' browsers that are now being developed or are currently deployed. These browsers are called 'next generation' because their main focus is to enable visualization of large amounts of data generated by 'NextGen' sequencing technologies. Two examples include the Anno-J browser (http://www.annoj.org) and JBrowse (42) (http://www.jbrowse.org). The main distinguishing characteristics between these browsers and the mainstream genome browsers reviewed above are how visualization is rendered and how the end-user interacts with the data. Both Anno-J and JBrowse use client-side technologies (e.g. AJAX, JavaScript) to render images rather than creating images on the server-side. By moving the computation from the server to the client, server load no longer impacts image rendering. The overall end-user experience tends to be smoother and more fluid because there are no page reloads and most requests happen in real-time. However, these browsers have a limited feature set when compared to other genome browsers. This can both be an advantage and a disadvantage. The advantage of having limited functionality is that it can handle large amounts of data very well. This browser ability will become increasingly important as more next-generation sequence data becomes available. Their major disadvantage is that they are largely untested. In addition, these browsers are limited in application platforms, availability of third-party plug-ins and the availability of tools for customization. Our survey results decidedly show that aside from visualization, maize cooperators want tools that facilitate their research, for example, tools that allow retrieval of data that is currently possible by the implementation of third-party plug-ins written by a community of developers.

It is important to note that compared to more established mainstream genome browsers, the next-gen browsers are still in early stages of development. With time, the data abundance generated by the next-generation sequencing technologies will push developers to tackle such challenges to create more mature client-side browsers, so that MODs can provide an improved service to their users.

Choosing a genome browser

Choosing a genome browser to address the maize community presents a challenge given that several browsers (reviewed above) have different strengths and weaknesses. For example, one of the most popular genome browsers, Ensembl, provides the best tools for comparative genomics. In contrast, another popular genome browser, GBrowse, provides a wide range of tools for MODs, yet its tool repertoire for comparative genomics is not as rich as Ensembl. Therefore, determining which software best suits the needs of maize geneticists is a task that requires a careful consideration.

Based upon results of the Genome Browser Survey, we chose GBrowse as the MaizeGDB Genome Browser for the following reasons:

(1) Because maize researchers have a wide range of research interests, we decided to implement a

genome browser that could be adapted to address general research questions. UCSC, xGDB, Ensembl and GBrowse would all fit this need.

(2) The UCSC genome browser is highly capable. However, one disadvantage of choosing it for the MaizeGDB Genome Browser would be that plant databases currently do not use the UCSC Genome Browser (an exception is the JGI 'tree of life', which uses the UCSC genome browser at http://genome.jgi-psf.org/ that also serves some plant genomes). TAIR, Soybase and SGN (among others) use GBrowse. The availability of developers from plant databases, as well as from other MODs (e.g. FlyBase and Mouse Genome Informatics), creates more opportunities for future collaboration to create similar solutions to respond to common challenges related to data integration and visualization.

(3) xGDB is a downloadable open source package, but it is not in wide use yet: so far PlantGDB is the only site that uses xGDB, and it has a limited number of developers.

(4) In the 'Feature ranking', the three most desired features are chosen as: ease of use, visuals and speed. The survey results indicate that cooperators do not consider Ensembl easy to use, and it is definitely perceived to be slow when compared to the other software available. Also, the desire to have cross-species comparison capability in a genome browser (where Ensembl excels) is only ranked 4. Note that although not currently as extensive as Ensembl, GBrowse has some cross-species tools already available [Synbrowse (43, 44), CMap (35) and GBrowse_syn, which is included in the GBrowse 1.70 Release].

(5) As indicated in the 'Indispensable features', cooperators would like to see specific tool development in a genome browser to enhance their research (e.g. finding genes between two markers). Therefore, a genome browser chosen by MaizeGDB should allow high flexibility in terms of code programming, tools development, and community involvement. The flexibility of tool development is intrinsic feature of GBrowse that allows customizable plug-in architecture as a community-based open source project. In the case of Ensembl, the code development is primarily done by a group in the UK and *ad hoc* tool development is carried out by research groups for their specific needs. Because this tool development by databases is specific to a particular Ensembl version, the tools must be modified or re-written for each new version of Ensembl. This creates an issue with Ensembl as it requires more manpower and funding to adapt the code to new version of the genome browser. In the case of xGDB, the flexibility in code development is somewhat limited. Because this browser is not widely used, the number of independent developers working on xGDB is not comparable to the community of GBrowse developers.

(6) MaizeSequence.org already provides maize genome sequence information using Ensembl. Providing this information using GBrowse and providing links to MaizeSequence.org would allow researchers to access different genome browsers for different applications and preferences. For example, when a cross-species comparison across many clades is necessary, Ensembl provides efficient solutions; however, when it comes to developing customizable visualization and analysis tools for maize-specific research problems, GBrowse stands out. Offering the availability of these two browsers to maize researchers will facilitate answering different research problems and will enhance agricultural research overall.

We realize that the accelerating technology would certainly engender new and improved genome browsers that are currently not available to be adopted and our current selection of a specific technology is likely to change as new technologies become available. That being said, at MaizeGDB, we are committed to being responsive to maize community needs and will remain open to adopting new technologies to address those needs.

Implementing GBrowse

We started implementing the GBrowse-based MaizeGDB Genome Browser (described in detail in ref. 10) in February 2008. We obtained maize data from various sources, including MaizeSequence.org, PlantGDB and MAGI. We chose five people for guidance (from academia and industry in the U.S. and abroad) and 10 people for beta testing among the cooperators who agreed at the end of the survey to be a part of the Genome Browser implementation. The guidance and beta-testing groups provided many valuable inputs to improve our users' experience with the MaizeGDB Genome Browser. The MaizeGDB Genome Browser was released in December 2008. We are still implementing ideas suggested by the guidance and beta testing groups and we continue to integrate the genome browser with existing data by creating novel tools and implementing existing tools as the needs to do so are identified. One of these suggestions, provided to us by Dr Sarah Hake, led to the creation and implementation of one of our most used tools in MaizeGDB: the Locus Lookup tool (45). This tool takes one or two loci as input and returns an approximate genomic region based on known physical and genetic associations, even in the case when the locus of interest is not yet placed on to the maize genome sequence. The utility of the Locus Lookup tool is apparent especially for the genomes that are in the process of being sequenced.

Supplementary Data

Supplementary data are available at *Database* Online.

Acknowledgements

We thank USDA-ARS for its sustained funding, past and present members of the MaizeGDB Working Group (Volker Brendel, Ed Buckler, Karen Cone, Mike Freeling, Owen Hoekenga, Anne-Francoise Lamblin, Thomas Lubberstedt, Karen McGinnis, Lukas Mueller, Mihai Pop, Marty Sachs, Pat Schnable, Tom Slezak, Anne Sylvester, and Doreen Ware), as well as the members of the MaizeGDB Executive Committee (Pat Schnable, Mary Alleman, Tom Brutnell, Sarah Hake, Jane Langdale, Jo Messing, Jean-Phillippe Vielle-Calzada, Anne Sylvester, William Tracy, Virginia Walbot, and Sue Wessler) for their direction, support, and inputs on this work. We also would like to thank the Genome Browser Guidance Group (Peter Balint-Kurti, Sarah Hake, Damon Lisch, Mike Muszynski, and Virginia Walbot) and Beta-testers (Alain Charcosset, Olivier Dugas, James Estill, David Hessel, Damon Lisch, Mike Muszynski, Paul Scott, Virginia Walbot, Rachel Wang, and Cesar Alvarez-Mejia), without whose comments and suggestions we could not have created an implementation of the MaizeGDB Genome Browser customized to support our users' needs. We thank the anonymous reviewers whose suggestions helped us improve this article. We very much appreciate the useful comments by Dr Patrick Armstrong. Last, but not least, we deeply appreciate and thank the maize community for their continuous support.

Funding

United States Department of Agriculture-Agricultural Research Service. Funding for open access charge: United States Department of Agriculture-Agricultural Research Service.

Conflict of interest. None declared.

References

1. Swarbreck,D., Wilks,C., Lamesch,P. *et al*. (2008) The Arabidopsis Information Resource (TAIR): gene structure and function annotation. *Nucleic Acids Res.*, **36**, D1009–D1014.
2. Tweedie,S., Ashburner,M., Falls,K. *et al*. (2009) FlyBase: enhancing Drosophila Gene Ontology annotations. *Nucleic Acids Res.*, **37**, D555–D559.
3. Liang,C., Jaiswal,P., Hebbard,C. *et al*. (2008) Gramene: a growing plant comparative genomics resource. *Nucleic Acids Res.*, **36**, D947–D953.
4. Mueller,L.A., Solow,T.H., Taylor,N. *et al*. (2005) The SOL Genomics Network: a comparative resource for Solanaceae biology and beyond. *Plant Physiol.*, **138**, 1310–1317.
5. Duvick,J., Fu,A., Muppirala,U. *et al*. (2008) PlantGDB: a resource for comparative plant genomics. *Nucleic Acids Res.*, **36**, D959–D965.
6. Childs,K.L., Hamilton,J.P., Zhu,W. *et al*. (2007) The TIGR Plant Transcript Assemblies database. *Nucleic Acids Res.*, **35**, D846–D851.
7. Chan,A.P., Pertea,G., Cheung,F. *et al*. (2006) The TIGR Maize Database. *Nucleic Acids Res.*, **34**, D771–D776.
8. Lawrence,C.J., Harper,L.C., Schaeffer,M.L. *et al*. (2008) MaizeGDB: The Maize Model Organism Database for basic, translational, and applied research. *Int. J. Plant Genomics*, **2008**, 496957.
9. Lawrence,C.J., Schaeffer,M.L., Seigfried,T.E. *et al*. (2007) MaizeGDB's new data types, resources and activities. *Nucleic Acids Res.*, **35**, D895–D900.
10. Sen,T.Z., Andorf,C.M., Schaeffer,M.L. *et al*. (2009) MaizeGDB becomes 'sequence-centric'. *Database*, doi:10.1093/database/bap020.
11. Schnable,P.S., Ware,D., Fulton,R.S. *et al*. (2009) The B73 maize genome: complexity, diversity, and dynamics. *Science*, **326**, 1112–1115.
12. Wei,F., Zhang,J., Zhou,S. *et al*. (2009) The physical and genetic framework of the maize B73 genome. *PLoS Genet.*, **5**, e1000715.
13. Vielle-Calzada,J.P., Martinez de la Vega,O., Hernandez-Guzman,G. *et al*. (2009) The Palomero genome suggests metal effects on domestication. *Science*, **326**, 1078.
14. Buckler,E.S., Holland,J.B., Bradbury,P.J. *et al*. (2009) The genetic architecture of maize flowering time. *Science*, **325**, 714–718.
15. McMullen,M.D., Kresovich,S., Villeda,H.S. *et al*. (2009) Genetic properties of the maize nested association mapping population. *Science*, **325**, 737–740.
16. Palmer,L.E., Rabinowicz,P.D., O'Shaughnessy,A.L. *et al*. (2003) Maize genome sequencing by methylation filtration. *Science*, **302**, 2115–2117.
17. Whitelaw,C.A., Barbazuk,W.B., Pertea,G. *et al*. (2003) Enrichment of gene-coding sequences in maize by genome filtration. *Science*, **302**, 2118–2120.
18. Lee,Y., Tsai,J., Sunkara,S. *et al*. (2005) The TIGR Gene Indices: clustering and assembling EST and known genes and integration with eukaryotic genomes. *Nucleic Acids Res.*, **33**, D71–D74.
19. Fu,H. and Dooner,H.K. (2002) Intraspecific violation of genetic colinearity and its implications in maize. *Proc. Natl Acad. Sci. USA*, **99**, 9573–9578.
20. Pruitt,K.D., Tatusova,T., Klimke,W. *et al*. (2009) NCBI Reference Sequences: current status, policy and new initiatives. *Nucleic Acids Res.*, **37**, D32–D36.
21. Grant,D., Nelson,R.T., Cannon,S.C. *et al*. (2009) SoyBase, The USDA-ARS Soybean Genome Database http://soybase.org.
22. Kurata,N. and Yamazaki,Y. (2006) Oryzabase. An integrated biological and genome information database for rice. *Plant Physiol.*, **140**, 12–17.
23. Kass,L.B., Bonneuil,C. and Coe,E.H. Jr. (2005) Cornfests, cornfabs and cooperation: the origins and beginnings of the maize genetics cooperation news letter. *Genetics*, **169**, 1787–1797.
24. Coe,E. (2009) East, Emerson, and the birth of maize genetics. In: Bennetzen,J. and Hake,S. (eds), *Handbook of Maize*. Springer, New York, NY, pp. 3–15.
25. Hubbard,T.J., Aken,B.L., Ayling,S. *et al*. (2009) Ensembl 2009. *Nucleic Acids Res.*, **37**, D690–D697.
26. Stein,L.D., Mungall,C., Shu,S. *et al*. (2002) The generic genome browser: a building block for a model organism system database. *Genome Res.*, **12**, 1599–1610.

27. Wolfsberg,T.G. (2007) Using the NCBI Map Viewer to browse genomic sequence data. *Curr. Protoc. Bioinformatics*, Chapter 1, Unit 15.
28. Karolchik,D., Kuhn,R.M., Baertsch,R. *et al*. (2008) The UCSC Genome Browser Database: 2008 update. *Nucleic Acids Res.*, **36**, D773–D779.
29. Schlueter,S.D., Wilkerson,M.D., Dong,Q. *et al*. (2006) xGDB: open-source computational infrastructure for the integrated evaluation and analysis of genome features. *Genome Biol.*, **7**, R111.
30. Paten,B., Herrero,J., Beal,K. *et al*. (2008) Enredo and Pecan: genome-wide mammalian consistency-based multiple alignment with paralogs. *Genome Res.*, **18**, 1814–1828.
31. Paten,B., Herrero,J., Fitzgerald,S. *et al*. (2008) Genome-wide nucleotide-level mammalian ancestor reconstruction. *Genome Res.*, **18**, 1829–1843.
32. Dowell,R.D., Jokerst,R.M., Day,A. *et al*. (2001) The distributed annotation system. *BMC Bioinformatics*, **2**, 7.
33. Jenkinson,A.M., Albrecht,M., Birney,E. *et al*. (2008) Integrating biological data - the Distributed Annotation System. *BMC Bioinformatics*, **9** (Suppl 8), S3.
34. Giardine,B., Riemer,C., Hardison,R.C. *et al*. (2005) Galaxy: a platform for interactive large-scale genome analysis. *Genome Res.*, **15**, 1451–1455.
35. Youens-Clark,K., Faga,B., Yap,I.V. *et al*. (2009) CMap 1.01: a comparative mapping application for the Internet. *Bioinformatics.*, **25**, 3040–3042.
36. Haider,S., Ballester,B., Smedley,D., Zhang,J., Rice,P. and Kasprzyk,A. (2009) BioMart Central Portal—unified access to biological data. *Nucleic Acids Res.*, **37**, W23–W27.
37. Muller,H.M., Kenny,E.E. and Sternberg,P.W. (2004) Textpresso: an ontology-based information retrieval and extraction system for biological literature. *PLoS Biol.*, **2**, e309.
38. Podicheti,R., Gollapudi,R. and Dong,Q. (2009) WebGBrowse—a web server for GBrowse. *Bioinformatics*, **25**, 1550–1551.
39. Rogers,A., Antoshechkin,I., Bieri,T. *et al*. (2008) WormBase 2007. *Nucleic Acids Res.*, **36**, D612–617.
40. Shaw,D.R. (2009) Searching the Mouse Genome Informatics (MGI) resources for information on mouse biology from genotype to phenotype. *Curr. Protoc. Bioinformatics*, Chapter 1, Unit 17.
41. Lander,E.S., Linton,L.M., Birren,B. *et al*. (2001) Initial sequencing and analysis of the human genome. *Nature*, **409**, 860–921.
42. Skinner,M.E., Uzilov,A.V., Stein,L.D. *et al*. (2009) JBrowse: a next-generation genome browser. *Genome Res.*, **19**, 1630–1638.
43. Brendel,V., Kurtz,S. and Pan,X. (2007) Visualization of syntenic relationships with SynBrowse. *Methods Mol. Biol.*, **396**, 153–163.
44. Pan,X., Stein,L. and Brendel,V. (2005) SynBrowse: a synteny browser for comparative sequence analysis. *Bioinformatics*, **21**, 3461–3468.
45. Andorf,C.M., Lawrence,C.J., Harper,L.C. *et al*. (2009) The Locus Lookup tool at MaizeGDB: identification of genomic regions in maize by integrating sequence information with physical and genetic maps. In: *Bioinformatics*, **26**, 434–436.