Published online 19 November 2007 *Nucleic Acids Research, 2008, Vol. 36, Database issue* ***D1041–D1043***
doi:10.1093/nar/gkm1022

Panzea: an update on new content and features

Payan Canaran[1], Edward S. Buckler[2,3], Jeffrey C. Glaubitz[4], Lincoln Stein[1], Qi Sun[5], Wei Zhao[1] and Doreen Ware[1,3,*]

[1]Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724, [2]Institute for Genomic Diversity, Cornell University, Ithaca, NY 14853-2703, [3]USDA-ARS NAA Plant, Soil & Nutrition Laboratory Research Unit, Tower Road, Ithaca, NY 14853-2901, [4]Laboratory of Genetics, University of Wisconsin, 425-G Henry Mall, Madison, WI, 53706 and [5]Computational Biology Service Unit, Cornell University, Ithaca, NY 14853, USA

Received September 24, 2007; Revised October 25, 2007; Accepted October 26, 2007

ABSTRACT

Panzea (http://www.panzea.org), the public web site of the project 'Molecular and Functional Diversity in the Maize Genome', has expanded over the past two years in data content, display tools and informational sections. The most significant data content expansions occurred for the single nucleotide polymorphism (SNP), sequencing, isozyme and phenotypic data types. We have enhanced our existing web display tools and have launched a number of new tools for data display and analysis. For example, we have implemented one that allows users to find polymorphisms between two accessions, a geographic map tool to visualize the geographic distribution of SNPs, simple sequence repeats (SSRs) and isozyme alleles and a graphical view of the placement of Panzea markers and genes/loci on genetic and physical maps. One goal of the informatics component of our project has been to generate code that can be used by other groups. We have enhanced our existing code base and have made our new tools available. Finally, we have also made available new informational sections as part of our educational and outreach efforts.

INTRODUCTION

The National Science Foundation-funded project 'Molecular and Functional Diversity in the Maize Genome' aims to examine the effect of selection on maize molecular diversity and how this molecular diversity relates to functional variation (1). The public web site of this project, Panzea (http://www.panzea.org), has expanded considerably over the past two years. We have continued to increase our molecular diversity data content by making new assay results available on our web site and via data dumps. We have enhanced our existing display tools and have implemented a number of new tools in order to better represent the underlying biology of our data. This project has generated a number of generic software packages that can be utilized by other groups. In the past two years, we have continued to make new software available and have also made improvements to our existing code base. In addition to expansion of our data and improvements to our software tools, we have made available a number of new informational sections as part of our educational and outreach efforts.

DATA EXPANSION

In the past two years, Panzea continued to expand its molecular diversity data content. The most significant data expansions occurred for the single nucleotide polymorphism (SNP), sequencing, isozyme and phenotypic data types. The number of SNP genotypic data points (two data points per diploid genotype) increased more than 9-fold, from 551 584 data points (in 985 marker assays) to 5 004 290 data points (in 4654 marker assays) between the August 2005 and September 2007 data releases. Within the same period, the number of individual sequences in sequence alignments almost doubled, from 67 703 (in 3542 alignments) to 130 033 (in 4696 alignments). The number of phenotypic data points increased almost 8-fold, from 34 124 to 262 150. Furthermore, 749 796 new isozyme genotypic data points were made available (plus an additional 201 263 isozyme allele frequency data points that are not displayed on the web site but are available in the database dumps), along with a number of smaller scale data sets. The number of simple sequence repeat (SSR) data points increased from 400 552 (in 612 marker assays) to 406 114 (in 636 assays) and 10 644 cleaved amplified polymorphic sequence (CAPS) marker data points (from four assays) and 26 126 indel marker data points (from eight assays) were added. The number of germplasm accessions made available on Panzea increased from 2790 to 2882. Detailed data statistics of the current data release can be accessed using the 'Stats' link on the navigation bar.

*To whom correspondence should be addressed. Tel: +1 516 367 6979; Fax: +1 516 367 6851; Email: ware@cshl.edu

NEW FEATURES

Panzea hosts a large quantity of molecular diversity data and a number of tools that allow users to query and display them. Since our previous NAR database issue publication (2), we have made a number of enhancements to our existing display tools. In addition, we have launched new display and analysis tools, as well as an intermediary page that simplifies linking to Panzea.

Marker-centric 'Molecular Diversity' page

The 'Molecular Diversity' page allows users to query the Panzea database to find our molecular markers and their available sequence alignments or genotypes. The initial implementation of the page allowed users to query for marker assays by gene/locus name and marker type (but not by marker name). The results were displayed in an assay-centric manner. Each assay was linked to additional pages that provided detailed views of the assay results. In order to better represent the underlying biology, we redesigned this page to be marker-centric. Users can now query for assays by marker name, in addition to previously available criteria, and the results are now displayed by marker: each marker in a result set links to results pages for all the assays performed for this marker. Furthermore, we have updated the detailed assay results display to reflect the new marker-centric grouping. In the current implementation, multiple assay results for the same marker are displayed in aggregate.

Polymorphism between accessions

We have implemented a new display that allows users to select an accession pair and view a comparison of their genotypes for the set of markers of a specified type that they have in common. All the genotypes at the common set of markers can be displayed, or the results can be filtered to show only the markers with genotypes that are polymorphic between the two accessions (or only those matching between the two accessions). The markers compared can also be restricted to those within a specified chromosomal region. This tool can be accessed using the 'Polymorphism between Accessions' link from the front page.

Geographic distribution of allele frequencies

In the results from a 'Molecular Diversity' search, the SNP and SSR assay data are displayed in a tabular format for each marker. Although this display provides a comprehensive view of the available data and allows further analysis via data dumps, it does not provide a view of the data points in a geographic context. In order to enable such a visualization, we implemented a geographic display that allows users to display the geographic distribution of allele frequencies for a given SNP or SSR marker on a geographic map. The display is interactive and allows users to move or zoom the map, as well as to filter data based on species/subspecies or for a particular allele. Users can access this display by searching for a marker on the 'Geographic Map Viewer' page and following 'Geo. Map' links to the geographic display.

Genetic and physical map viewer

We implemented a graphical view of genetic and physical maps of Panzea markers and genes/loci using the CMap Comparative Map Viewer (http://www.gmod.org/cmap). The placement of our markers can be graphically viewed on two genetic maps and one physical map. The two genetic maps are the IBM2 2005 Neighbors map (3) and a map of a maize-teosinte backcross population (4). The physical map is the Maize Agarose FPC (fingerprint contig) map of the Arizona Genomics Institute (5). We developed a custom front end to the standard CMap interface, which can be accessed using the 'Genetic and Physical Map Viewer' link from the front page. Our front end allows simplified access to the underlying views while retaining the full power of CMap, by launching it at the final step of the selection.

Linking to Panzea

Panzea provides a number of pages to access the underlying data. Linking to each page requires provision of a number of parameters in a URL. We have implemented an intermediary gateway page that allows external developers to link to Panzea pages using a standard and simple URL syntax. The gateway page recognizes the standard parameters and redirects the link to the appropriate search page. The page can be updated transparently to account for changes made to individual pages. Therefore, this indirect linking structure allows pointing to Panzea pages using stable URLs, which are not affected by changes to individual pages. More information on this utility is available from the 'Search Page and File Gateway' link on the front page.

Code base

One goal of the bioinformatics component of this project was to generate generic, well-documented software tools that can be used by other groups. The tools we developed previously were made publicly available. In the past two years, we repackaged our existing tools and incorporated automatic installation methods, significantly simplifying their installation and customization. We have improved and added features to the code base in response to external requests, facilitating their use. In addition, we packaged and made available the new tools that we have developed. A number of these packages are available through the Comprehensive Perl Archive Network (CPAN) (http://search.cpan.org) as standard Perl distributions that can be easily installed using standard Perl installation tools, such as the CPAN shell. The remaining code base is available as stand-alone packages on the Panzea web site. Currently, the code base is available as three Perl distributions, HTML::SearchPage (a generic framework for building web-based search pages), HTML::GMap (a high-level Perl wrapper around the Google Maps™ application programming interface) and GD::Icons (a supplementary module for generating icons), and two stand-alone software packages, Look-Align (6) (a generic stand-alone web-based alignment viewer) and Panzea-Searches (a Panzea-specific code base package).

The 'Download Site Software' link on the front page provides detailed information on the code base.

Other

We have made a number of new pages available as part of our educational and outreach efforts. We have added a section on the project significance and a page with links to a number of relevant sites. We have also made the Maize Domestication and the History of Maize education slides available in Spanish. These pages and files are accessible through the 'Project Significance', 'Links' and 'Educational Resources' links on the front page. In the results of a Gene/Locus search, links are provided from contigs on the FPC agarose map to their detailed map at the Arizona Genomics Institute site and from a maize chromosome to the IBM2 2005 neighbors map of that chromosome at MaizeGDB (http://www.maizegdb.org/) (7). In addition, the sequence trace files generated by the sequencing assays are now available for download from the web site, accessible from the 'Data Sets' link on the front page. In addition, we have made the context sequences for SNP assays and reference sequences for sequencing assays available by incorporating them into the 'Molecular Diversity' search results. Finally, we have implemented a page that allows users to search a number of sequence databases, including the Panzea maize reference sequences (comprised of one representative sequence from each sequence alignment), using the wu-blast (http://blast.wustl.edu) application. This utility can be accessed from the 'BLAST' link on the front page.

Future plans

Our future plans for Panzea include accurate placement of genes, loci, markers and QTL in the context of the nascent genome sequence for maize inbred B73 (http://www.maizesequence.org/) and development of tools for tabular and graphical display of QTL mapping results. Beyond the end of the Molecular and Functional Diversity in the Maize Genome project, we anticipate that the Panzea site will be maintained for as long as possible by leveraging resources from other projects. Furthermore, the data content of Panzea, and much of the value that we have added to the data, will be made available via MaizeGDB and via the Germplasm Repository and Information Network (GRIN) of the US Department of Agriculture (http://www.ars-grin.gov/).

ACKNOWLEDGEMENTS

We thank all members of the Molecular and Functional Diversity in the Maize Genome project for providing data, educational materials and/or technical support for Panzea. This work was funded by National Science Foundation Plant Genome Grant, Division of Biological Infrastructure (0321467); US Department of Agriculture-Agricultural Research Service. Funding to pay the Open Access publication charges for this article was provided by National Science Foundation (0321467).

Conflict of interest statement. None declared.

REFERENCES

1. Buckler,E.S., Gaut,B.S. and McMullen,M.D. (2006) Molecular and functional diversity of maize. *Curr. Opin. Plant Biol.*, **9**, 172–176.
2. Zhao,W., Canaran,P., Jurkuta,R., Fulton,T., Glaubitz,J., Buckler,E., Doebley,J., Gaut,B., Goodman,M. *et al.* (2006) Panzea: a database and resource for molecular and functional diversity in the maize genome. *Nucleic Acids Res.*, **34** (Database issue), D752–D757.
3. Schaeffer,M.L., Sanchez-Villeda,H., McMullen,M.D. and Coe,E.H. Jr (2006) IBM2 2005 Neighbors—45,000 locus resource for maize. Plant and Animal Genome XIV Conference. P372. http://www.intl-pag.org/14/abstracts/PAG14_P372.html.
4. Briggs,W.H., McMullen,M.D., Gaut,B.S. and Doebley,J. (2007) Linkage mapping of domestication loci in a large Maize-Teosinte Backcross Resource. *Genetics,* doi: 10.1534/genetics.107.076497.
5. Gardiner,J., Schroeder,S., Polacco,M.L., Sanchez-Villeda,H., Fang,Z., Morgante,M., Landewe,T., Fengler,K., Useche,F. *et al.* (2004) Anchoring 9,3971 maize expressed sequence tagged unigenes to the bacterial artificial chromosome contig map by two-dimensional overgo hybridization. *Plant Physiol.*, **134**, 1317–1326.
6. Canaran,P., Stein,L. and Ware,D. (2006) Look-Align: an interactive web-based multiple sequence alignment viewer with polymorphism analysis support. *Bioinformatics*, **22**, 885–886.
7. Lawrence,C.J., Schaeffer,M.L., Seigfried,T.E., Campbell,D.A. and Harper,L.C. (2007) MaizeGDB's new data types, resources and activities. *Nucleic Acids Res.*, **35** (Database Issue), D895–D900.