

Panzea: a database and resource for molecular and functional diversity in the maize genome

Wei Zhao¹, Payan Canaran¹, Rebecca Jurkuta², Theresa Fulton³, Jeffrey Glaubitz², Edward Buckler^{3,4}, John Doebley², Brandon Gaut⁵, Major Goodman⁶, Jim Holland^{6,7}, Stephen Kresovich³, Michael McMullen^{8,9}, Lincoln Stein¹ and Doreen Ware^{1,4,*}

¹Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724, USA, ²Genetics Department, University of Wisconsin-Madison, Madison, WI 53706, USA, ³Institute for Genomic Diversity, Cornell University, Ithaca, NY 14853-2703, USA, ⁴USDA-ARS NAA Plant, Soil & Nutrition Laboratory Research Unit, Tower Road, Ithaca, NY 14853-2901, USA, ⁵Department of Ecology and Evolutionary Biology, University of California-Irvine, 92697-2525, USA, ⁶Crop Science Department, North Carolina State University, Raleigh, NC 27695-7620, USA, ⁷USDA-ARS, Plant Science Research Unit, Raleigh, NC 27695-7620, USA, ⁸Department of Agronomy, University of Missouri-Columbia, Columbia, MO 65211-7020, USA and ⁹USDA-ARS, Plant Genetics Research Unit, Columbia, MO 65211, USA

Received August 10, 2005; Revised September 1, 2005; Accepted September 12, 2005

ABSTRACT

Serving as a community resource, Panzea (<http://www.panzea.org>) is the bioinformatics arm of the Molecular and Functional Diversity in the Maize Genome project. Maize, a classical model for genetic studies, is an important crop species and also the most diverse crop species known. On average, two randomly chosen maize lines have one single-nucleotide polymorphism every ~100 bp; this divergence is roughly equivalent to the differences between humans and chimpanzees. This exceptional genotypic diversity underlies the phenotypic diversity maize needs to be cultivated in a wide range of environments. The Molecular and Functional Diversity in the Maize Genome project aims to understand how selection has shaped molecular diversity in maize and then relate molecular diversity to functional phenotypic variation. The project will screen 4000 loci for the signature of selection and create a wide range of maize and maize-teosinte mapping populations. These populations will be genotyped and phenotyped, permitting high-power and high-resolution dissection of the traits and relating the molecular diversity to functional variation. Panzea provides access to the genotype, phenotype and polymorphism data produced by the project through user-friendly web-based database searches

and data retrieval/visualization tools, as well as a wide variety of information and services related to maize diversity.

INTRODUCTION

Maize (*Zea mays* ssp. *mays*) is one of the most important crop species in the world, feeding both the world's people and livestock, while also providing necessary raw materials for several industries (1). Maize is also the most diverse crop species known, containing tremendous variation in morphological traits and extensive polymorphism in DNA sequences (2). On average, two randomly chosen modern maize lines have one single-nucleotide polymorphism (SNP) every ~100 bp (3); despite being a single interbreeding sub-species, this divergence is roughly equivalent to the differences between humans and chimpanzees (4,5). This exceptional genotypic diversity underlies the phenotypic diversity maize needs to be cultivated in a wide range of environments, ranging from deserts to tropical rainforest, to high mountains and to short growing seasons in Canada. Understanding the mechanisms underlying the morphological and genetic diversity that exists in maize is critical to future plant breeding and is of great interest to maize geneticists and plant breeders.

Studies have provided both genetic and molecular evidence to indicate that maize was domesticated ~9000 years ago from a form of teosinte, known as *Zea mays* ssp. *parviglumis* or Balsas teosinte (2,6), which grows commonly as a wild plant in the valleys of southwestern Mexico. Maize and teosinte

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

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors

© The Author 2006. Published by Oxford University Press. All rights reserved.

The online version of this article has been published under an open access model. Users are entitled to use, reproduce, disseminate, or display the open access version of this article for non-commercial purposes provided that: the original authorship is properly and fully attributed; the Journal and Oxford University Press are attributed as the original place of publication with the correct citation details given; if an article is subsequently reproduced or disseminated not in its entirety but only in part or as a derivative work this must be clearly indicated. For commercial re-use, please contact journals.permissions@oxfordjournals.org

exhibit striking differences in their adult morphologies, and the domestication of maize has resulted in highly modified inflorescence and plant architecture (7). Like most domesticated plants and animals, maize had experienced a 'domestication bottleneck' that reduced its genetic diversity relative to its wild ancestor teosinte (8). This bottleneck was a consequence of the limited pool of wild founder plants from which domesticated maize arose, and its effects were genome wide. In addition, a 'selection bottleneck' of greater severity was experienced by specific genes that were the direct targets of artificial selection, both during domestication and later, during the process of plant improvement; the reduction of genetic diversity in these genes was above and beyond that caused by the domestication bottleneck (9). Genes that underwent selection are probably those of agronomic importance.

'The Molecular and Functional Diversity in the Maize Genome' project (referred to hereafter as 'the Maize Diversity project') aims to understand how natural and artificial selection has shaped molecular diversity in maize and to relate molecular diversity to functional phenotypic variation. To address the first question—How has selection shaped molecular diversity?—SNP discovery has been performed on 1000 genes using diverse maize inbred lines (including tropical/temperate inbreds) and teosinte lines. Analysis in 774 genes indicates that 2–4% of these genes experienced artificial selection (9). In addition, a range of tests of selection will be performed on a total of 4000 genes to identify genes that show evidence for positive, diversifying and purifying selection. The identified genes will be those involved in the process of domestication, agronomic improvement and local adaptation. To address the second question—How does this molecular diversity relate to functional trait variation?—the project will be creating a wide range of maize and maize–teosinte linkage and association mapping populations that will capture a tremendous range of diversity. These populations will be genotyped for SNPs in genes that putatively affect traits of interest, and phenotyped for these traits. Together, these studies will permit high-power and high-resolution dissection of the traits, and thus will relate the molecular diversity to functional variation.

As the bioinformatics arm of the project and a community resource for molecular and functional diversity in the maize genome, Panzea (<http://www.panzea.org>), provides researchers worldwide and the public with access to the genotype, phenotype and polymorphism data produced by the project. In addition, Panzea also provides an assortment of information and services related to the topic of maize diversity, including information on software for association studies and statistical genetics analyses, various downloadable datasets, maize diversity literature and educational resources.

DATABASE SEARCHES AND TOOLS

Panzea contains the genotype, phenotype and polymorphism data produced by the Maize Diversity project. As of August 2005, 222 simple sequence repeat (SSR) loci were scored on 1543 accessions for a total of 400 552 data points, 604 SNP loci were scored on 282 accessions for a total of 551 584 data points and 3143 loci were sequenced from 177 accessions for a total of 67 703 data points. In addition, there are 34 124

phenotype experiment data points. User-friendly web-based database searches, data retrieval tools and data visualization tools enable access to data, which can be viewed on the web and/or downloaded as ASCII text file, tab-delimited text file or Microsoft® Excel spreadsheet. All database search interfaces also allow the user to sort search results by columns of interest. Currently the following database searches and tools are available.

Germplasm search

Germplasm refers to a collection of biological material. For maize, germplasm is typically available in the form of seed samples from particular accessions stored in germplasm banks, such as stock centers. An accession is an entry in a germplasm bank for a collection of seed from a particular inbred line, landrace or from a particular plant, family or population. The germplasm pool used by the project currently contains 2790 accessions, including 1567 open pollinated maize populations, 430 maize inbred lines, 296 maize hybrid, 468 teosinte populations, 23 teosinte inbred lines and 6 *Tripsacum* samples. The germplasm search allows the user to search the germplasm data and provides detailed information on germplasm type (inbred, hybrid, open pollinated, teosinte and so on), country of origin, geographical parameters of the collection location of the accession and the accession source (where the seeds can be obtained).

Gene/Locus search

The gene/locus search allows the user to browse or search all genes and loci within the project's data collection. Here 'gene' is used to refer only to loci that have been genetically characterized and given a gene name (e.g. *tb1*) by maize geneticists. 'Loci' are of several different types, including SNP, SSR, genomic or cDNA clones, ESTs and cytological loci. The information available about each gene/locus entry includes name and type of the gene/locus, genetic position within the maize genome and location on the maize Agarose FPC physical map (10) whenever these data are available, and comments added by members of the project. In addition, if assays exist for polymorphisms within the gene/locus, links to detailed information on length assays (SSRs, AFLPs, indels), SNP assays or sequence assays (sequence alignments for SNP discovery) are also provided.

Molecular diversity search

This search allows the user to browse or search data related to the project's genotyping experiments. Search results provide detailed information on each experiment, including the gene/locus, the type of markers used (SNP, SSR, CAPS, INDEL and sequencing), as well as primer sequences and repeat motif for SSR markers. Moreover, this page also provides links to detailed SSR and SNP assay results whenever such assay data are available. Currently there are a total of 5183 molecular diversity records in Panzea.

Phenotype search

The phenotype search allows the user to browse or search the phenotypic data gathered in the course of the project. These data are collected for a variety of accessions and will be used for Quantitative trait locus (QTL) and association studies.

Currently there are 34 239 phenotype records in the database; most records contain plot means for a given accession. The information in each record includes background and source of the accession and its growth circumstances (evaluation locality, state, country and planting date), as well as the phenotype name and value. Links to the contact information of the seed source also appear, subject to availability.

Genomic Diversity and Phenotype Connection (GDPC) data browser

The GDPC is a middleware database interface that provides access to integrated data on genomic diversity in a standard format (11). The GDPC interface consists of three principal elements: behind-the-scenes connections to databases, the GDPC java API and the GDPC data browser. The GDPC browser available at Panzea allows the user to retrieve data from the Panzea database and other 'GDPC-enabled' data sources simultaneously and to analyze the integrated data. Data accessible through the GDPC browser include phenotypic data and genomic diversity data, such as SNP and SSR genotypes and sequences.

Alignment viewer

The alignment viewer displays the pre-computed multiple sequence alignment data that the project generates. Users can query for alignments using either assay identifiers or gene/locus names. For each alignment, the viewer provides an overview containing points of variation as well as detailed alignment views at the level of the base pair. The detailed alignment view displays consensus sequence, highlighted variation points and related statistics. To allow the user to remove low-quality sequences or segments and sequences that do not align well with the remainder, an on-the-fly filtering procedure accompanies the viewer; filtering is user-configurable through the web interface.

RESOURCES

Gene suggestion page

The Maize Diversity project will be conducting association studies with a large number of genes for agronomic and developmental traits. Diverse samples of maize and teosinte will be genotyped using SNPs discovered within genes and phenotyped for traits of interest. Association studies will be subsequently conducted to find links between the genes and the traits of interest. The maize research community is invited to make suggestions for genes that will be examined in the association studies. Suggestions can be made using the gene suggestion page (<http://www.panzea.org/sug/index.html>).

Germplasm development at Panzea

The Maize Diversity project has been developing germplasm resources that it will make available to the community as they become available. The germplasm that are being developed include 12 teosinte inbred lines, 1000 maize-teosinte recombinant inbred lines (RILs), 7000 maize RILs (25 populations of >200 lines each, derived from crosses between the elite line B73 and 25 diverse maize lines) and a maize half-diallel (pairwise crosses between the 25 diverse lines used to generate the RILs). Detailed information and the anticipated release

date of each resource can be found at <http://www.panzea.org/lit/germplasm.html>. In addition, 281 diverse maize inbred lines from throughout the world that were genotyped with SSRs in the previous phase of the project (12) are currently available through the North Central Regional Plant Introduction Station; a list of the lines can be downloaded from the Panzea website.

Educational resources

As one of the best examples of crop domestication and an excellent visual example of genetic inheritance, maize is an exceptional educational tool that could be used in an integrated approach to education. Maize offers a good demonstration of the plant life cycle owing to its large size, and it has played an important role throughout history and in many cultures, most particularly that of Native American and Meso-American peoples (13). The Maize Diversity project has developed mobile 'story boards' that explain the domestication of maize, from perspectives of both history and genetics. These story boards, introduced in person and then exhibited in school hallway display cases, will help to bring the many-faceted story of maize to students in very rural areas or high-minority-resource-poor areas, where ready access to computers is not available. Slide presentations have been developed to introduce the 'story boards'; they can be viewed and downloaded at the Panzea website.

Maize evolution and diversity literature

Maize evolution and diversity has been an active area of research for more than a century. The Maize Diversity project has assembled a bibliography for much of the literature on this topic from the 19th century to the present and this is presented at the Panzea website. Users can browse the bibliography either chronologically by publication date or alphabetically by first author. Links to the articles are provided where possible.

USAGE CASE

A practical usage case will illustrate how to access information available through Panzea. Imagine that a researcher has an interest in the variation within and surrounding a particular maize gene, e.g. *teosinte branched1* (*tb1*), a well-known maize gene with an important role in domestication (7). The researcher might wish to perform a fine-scale QTL analysis focusing on this genomic region (14), or perhaps examine the genomic extent of the selective sweep associated with this locus (15). The goal of the researcher is thus to find available SNP and microsatellite markers, both within the *tb1* gene and in its surrounding genomic region.

To accomplish this goal, the researcher first performs a gene/locus search for the gene/locus 'tb1', using the search operator 'equals'. The results of this search will indicate that *tb1* is on chromosome 1 at position 691.6 (in centiMorgans) on the 'IBM2 2004 Neighbors' genetic map (for details on this map see <http://www.maizemap.org>). Another gene/locus search is then performed for all mapped loci on chromosome 1, with the results sorted by 'IBM2 Position', then by 'Genetic Bin', and finally by 'FPC contig'. The researcher can then navigate through the search results to the page containing

Gene/Locus Search

From this search utility you can download information about all the loci used in this project, including name and type of the loci as well as information about their location within the maize genome. Types of loci include Mapped, Unmapped, and cytological loci.

To obtain this information, use the search form at the right. To obtain a complete list of the loci used in our project, leave all the search parameters in their default state, choose the output format you would like, and click Submit. To limit the search by your criteria of interest, use the drop-down menus to specify your criteria before clicking Submit.

The text input field (Accession) accepts '*' and '?' as wildcards. '?' stands for a single character and '*' stands for zero or more characters. The search is not case-sensitive.

Output tables in HTML format will appear below the search form on a refreshed page. Output tables in ASCII format will appear in a new window. **To save the output table to your computer**, select Excel as your output format, submit your query, select "Save file as..." in the dialog pop-up window, and use ".xls" for the file extension.

[View or Download](#) a Gene / Locus Search tutorial

Gene/Locus:

Locus Type:

IBM2 Chromosome:

Sort by (#1):

Sort by (#2):

Sort by (#3):

Output Format:

Record 51-100 of 141

Page 2 of 3

Page 2 Go

| Name | Locus Type | Unigene Name | Assays | IBM2 Chr. | IBM2 Position | Genetic Bin | FPC Contig | FPC Chr. | FPC Start | FPC Stop | Other FPC Contigs |
|----------|--------------|--------------|---|-----------|---------------|-------------|------------|----------|-----------|----------|-------------------|
| umc1706 | Mapped locus | | [SSR:5548] | 1 | 670.59 | 1.07 | ctg38 | 1 | 718 | 736 | |
| pgm1 | Mapped locus | | | 1 | 688.8 | 1.09 | | | | | |
| tb1 | Mapped locus | | [CAPS:4573] [CAPS:4574] [CAPS:4575] [INDEL:4576] [INDEL:4577] [INDEL:4578] [INDEL:4579] [Sequencing:270] [Sequencing:334] [SNP:3614] [SNP:3615] [SNP:3616] [SNP:3617] [SNP:3618] [SNP:3619] [SNP:3620] [SNP:3621] [SNP:3622] [SNP:3623] [SNP:3624] [SNP:3625] [SNP:3626] [SNP:3627] [SNP:3628] [SNP:3629] [SNP:3630] [SNP:3631] [SNP:3632] [SNP:3633] [SNP:4587] [SNP:4588] [SNP:4589] [SNP:4590] [SNP:4591] [SNP:4592] [SNP:4593] [SNP:4594] [SNP:4595] [SNP:4596] [SNP:4597] | 1 | 691.6 | 1.09 | | | | | |
| d8 | Mapped locus | | [Sequencing:247] [Sequencing:332] [SNP:3599] [SNP:3600] [SNP:3601] | 1 | 694.75 | 1.09 | | | | | |
| AY110356 | Mapped locus | CL22447_1 | [Sequencing:2757] | 1 | 706.4 | | conflict | | | | ctg361, ctg40 |
| phi002 | Mapped locus | | [SSR:5653] | 1 | 720.9 | 1.08 | conflict | | | | ctg40, ctg96 |
| AY110159 | Mapped locus | CL31053_1 | [Sequencing:2826] | 1 | 722.3 | | | | | | |
| phi037 | Mapped locus | | [SSR:5312] [SSR:5345] | 1 | 722.4 | 1.08 | ctg40 | 1 | 930 | 947 | |

Figure 1. Use of a gene/locus search to find Panzea markers both within and flanking a gene of interest (e.g. *tb1*) on the IBM2 2004 Neighbors genetic map of maize.

the *tb1* region of the IBM2 2004 Neighbors map (Figure 1). Here it can be seen (in the 'Assays' column) that we have identified 31 SNPs at the *tb1* locus, and that 20 of these have links to 'Assay Results', indicating that they are validated SNPs for which Panzea has genotypic data. Researchers can develop their own assays for one or more of these validated

SNPs based upon their 'context sequences'. The context sequences are derived from the consensus sequence of our SNP discovery panel, and show the target SNP enclosed within square brackets, and other flanking polymorphisms in curly brackets. Users can download Microsoft® Excel files containing the context sequences for all of our validated

SNPs from the 'Datasets' page in the 'Resources' section of the Panzea site. In a similar fashion, the researchers can develop their own assays for validated SNPs from loci flanking *tb1* (e.g. *d8*). In the future, Panzea will have a direct link to the context sequence for a particular SNP from the molecular diversity search results, which in turn will be linked to from the gene/locus search results.

To facilitate the use of Panzea microsatellite (SSR) markers, Panzea provides PCR primer sequences. These can be found via the molecular diversity search. For example, assume that the researcher wishes to use the SSR marker 'umc1706', shown by the above gene/locus search to be near the *tb1* locus. A molecular diversity search for 'gene/locus equals umc1706' will yield information on the marker assay, including the sequences of the forward and reverse PCR primers. A similar strategy can reveal other flanking microsatellite markers identified through the gene/locus search.

Since the gene/locus search uses map information not only from the IBM2 2004 Neighbors genetic map of maize, but also from the Agarose FPC physical map (10), the researcher might be able to find additional loci near *tb1* (or another gene of interest) on the sole basis of physical position. Results from our gene/locus search of chromosome 1 (Figure 1) indicate that *tb1* is in the vicinity of FPC contigs 38–40. After an additional gene/locus search of loci with a 'NULL' value for (IBM2) chromosome, and with the results sorted by 'FPC Chromosome', by 'FPC Contig' and finally by 'FPC Start', the researcher can navigate to the page (or pages) showing loci that are on FPC contigs 38–40 but not on the IBM2 genetic map. For example, there is a validated SNP marker for locus AY106760 on contig 40. Following the link to 'Assay Results' for this SNP marker reveals that the marker name is 'PZA00658.19'. The context sequence for this SNP can then be found in the Excel file on the 'Datasets' page.

DATABASE AND WEBSITE IMPLEMENTATION

The Panzea website is hosted on a Dell PowerEdge 2650 server running Linux. The back end of the web site is a MySQL database, and the front end consists of a set of Perl CGI scripts and modules, running under a mod_perl-enabled Apache web server.

The Panzea database schema is based on the Genomic Diversity and Phenotype Data Model (GDPDM) (<http://www.maizegenetics.net/gdpdm/>). The schema and the table and field descriptions are available at the Panzea web site.

AVAILABILITY

All Perl CGI scripts and modules used as the front end of the Panzea web site are open-source software, and can be downloaded at the Panzea web site.

All data contained in the Panzea database are freely available to the public. Users can download the entire Panzea database or partial datasets at the Panzea web site. The entire database is available as a MySQL database dump, while the partial datasets, including passport (germplasm) data, phenotype data, sequence data, SNP data and SSR data, are available as tab-delimited text files.

Presently, maize diversity data exist in multiple repositories, and researchers must go to various places to find all available data. To help to make this problem less troubling, Panzea will make its data available for any reference databases to mirror and will establish reciprocal links with those databases. It is expected that Panzea data will be integrated with other diversity data in one or more long-term data repositories, such as Gramene (16), MaizeGDB (17) and Germplasm Resources Information Network (GRIN).

FUTURE ENHANCEMENTS

As more datasets and analysis tools become available, the implementation of display pages will mature over the course of the project. The immediate aim is to develop user-friendly customized queries and advanced data-browsing tools in the near future. In addition, we also plan to add the following major new elements to Panzea.

Gene annotation

These annotations will provide information on the genes, including sequence, genetic and physical map position, diversity statistics, associated maize and rice sequences, associated plant anatomy terms derived from the EST libraries, and mappings to maize and rice genomic sequences. Protein annotations will include mappings to maize UniProt (18) proteins, Pfam (19) and PROSITE (20) mappings, and InterPro (21) to gene ontology (GO) (22) mapping. These annotations will provide key information for classifying selection effects and potential trait associations.

Germplasm displays

The germplasm used in this project and associated genotypic and phenotypic data will appear in both geographical and phylogenetic tree displays. These displays will help dissect the ecological distribution of the landraces. The germplasm pages will also provide external links to the GRIN (<http://www.ars-grin.gov/>) for additional information on germplasm accessions.

QTL display/analysis

Tools are in development to display the results of QTL detection experiments using an integrated approach that will provide QTL positions across multiple populations, markers within intervals, sequence diversity and gene associations. Plant Ontologies (23) and Trait Ontologies (24) will help organize the display of QTL for related traits. Genes within QTL intervals will be made searchable by GO terms to facilitate the identification of genes, based on molecular function, biological process and cellular location. Where significant collinear regions with rice are present, there will be links to the rice QTL and sequence-based maps. In addition, tools are being developed to facilitate running bulk QTL analyses.

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 is supported by NSF Grant 0321467 and USDA ARS. Funding to pay the Open Access publication charges for this article was provided by NSF Grant 0321467.

Conflict of interest statement. None declared.

REFERENCES

- Fussell, B. (1999) *The Story of Corn*. 2nd edn. North Point Press, NY.
- Matsuoka, Y., Vigouroux, Y., Goodman, M.M., Sanchez, G.J., Buckler, E. and Doebley, J. (2002) A single domestication for maize shown by multilocus microsatellite genotyping. *Proc. Natl Acad. Sci. USA*, **99**, 6080–6084.
- Tenaillon, M.I., Sawkins, M.C., Long, A.D., Gaut, R.L., Doebley, J.F. and Gaut, B.S. (2001) Patterns of DNA sequence polymorphism along chromosome 1 of maize (*Zea mays* ssp. *mays* L.). *Proc. Natl Acad. Sci. USA*, **98**, 9161–9166.
- Chen, F.C. and Li, W.H. (2001) Genomic divergences between humans and other hominoids and the effective population size of the common ancestor of humans and chimpanzees. *Am. J. Hum. Genet.*, **68**, 444–456.
- Buckler, E.S. and Stevens, N.M. (2006) Maize Origins, Domestication, and Selection. *Darwin's Harvest*. Columbia University Press, NY.
- Doebley, J. (1990) Molecular evidence and the evolution of maize. *Econ. Bot.*, **44**, 6–27.
- Doebley, J. (2004) The genetics of maize evolution. *Annu. Rev. Genet.*, **38**, 37–59.
- Eyre-Walker, A., Gaut, R.L., Hilton, H., Feldman, D.L. and Gaut, B.S. (1998) Investigation of the bottleneck leading to the domestication of maize. *Proc. Natl Acad. Sci. USA*, **95**, 4441–4446.
- Wright, S.I., Bi, I.V., Schroeder, S.G., Yamasaki, M., Doebley, J.F., McMullen, M.D. and Gaut, B.S. (2005) The effects of artificial selection on the maize genome. *Science*, **308**, 1310–1314.
- Coe, E., Cone, K., McMullen, M., Chen, S.S., Davis, G., Gardiner, J., Liscum, E., Polacco, M., Paterson, A., Sanchez-Villeda, H. *et al.* (2002) Access to the maize genome: an integrated physical and genetic map. *Plant Physiol.*, **128**, 9–12.
- Casstevens, T.M. and Buckler, E.S. (2004) GDPC: connecting researchers with multiple integrated data sources. *Bioinformatics*, **20**, 2839–2840.
- Liu, K., Goodman, M., Muse, S., Smith, J.S., Buckler, E. and Doebley, J. (2003) Genetic structure and diversity among maize inbred lines as inferred from DNA microsatellites. *Genetics*, **165**, 2117–2128.
- Viola, H.J. and Margolis, C. (1991) *Seeds of Change: A Quincentennial Commemoration*. Smithsonian Institution Press, Washington, D.C.
- Doebley, J., Stec, A. and Gustus, C. (1995) *teosinte branched1* and the origin of maize: evidence for epistasis and the evolution of dominance. *Genetics*, **141**, 333–346.
- Clark, R.M., Linton, E., Messing, J. and Doebley, J.F. (2004) Patterns of diversity in the genomic region near the maize domestication gene *tb1*. *Proc. Natl Acad. Sci. USA*, **101**, 700–707.
- Ware, D., Jaiswal, P., Ni, J., Pan, X., Chang, K., Clark, K., Teytelman, L., Schmidt, S., Zhao, W., Cartinhour, S. *et al.* (2002) Gramene: a resource for comparative grass genomics. *Nucleic Acids Res.*, **30**, 103–105.
- Lawrence, C.J., Dong, Q., Polacco, M.L., Seigfried, T.E. and Brendel, V. (2004) MaizeGDB, the community database for maize genetics and genomics. *Nucleic Acids Res.*, **32**, D393–D397.
- Bairoch, A., Apweiler, R., Wu, C.H., Barker, W.C., Boeckmann, B., Ferro, S., Gasteiger, E., Huang, H., Lopez, R., Magrane, M. *et al.* (2005) The Universal Protein Resource (UniProt). *Nucleic Acids Res.*, **33**, D154–D159.
- Bateman, A., Coin, L., Durbin, R., Finn, R.D., Hollich, V., Griffiths-Jones, S., Khanna, A., Marshall, M., Moxon, S., Sonnhammer, E.L. *et al.* (2004) The Pfam protein families database. *Nucleic Acids Res.*, **32**, D138–D141.
- Hulo, N., Sigrist, C.J., Le Saux, V., Langendijk-Genevaux, P.S., Bordoli, L., Gattiker, A., De Castro, E., Bucher, P. and Bairoch, A. (2004) Recent improvements to the PROSITE database. *Nucleic Acids Res.*, **32**, D134–D137.
- Mulder, N.J., Apweiler, R., Attwood, T.K., Bairoch, A., Bateman, A., Binns, D., Bradley, P., Bork, P., Bucher, P., Cerutti, L. *et al.* (2005) InterPro, progress and status in 2005. *Nucleic Acids Res.*, **33**, D201–D205.
- Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T. *et al.* (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nature Genet.*, **25**, 25–29.
- The Plant Ontology™ Consortium. (2002), The Plant Ontology™ Consortium and Plant Ontologies. *Comp. Funct. Genomics*, **3**, 137–142.
- Jaiswal, P., Ware, D., Ni, J., Chang, K., Zhao, W., Schmidt, S., Pan, X., Clark, K., Teytelman, L., Cartinhour, S. *et al.* (2002) Gramene: development and integration of trait and gene ontologies for rice. *Comp. Funct. Genomics*, **3**, 132–136.