Nucleic Acids Research, 1999, Vol. 27, No. 1 95–98

The Mouse Genome Database (MGD): genetic and genomic information about the laboratory mouse

Judith A. Blake*, Joel E. Richardson, Muriel T. Davisson, Janan T. Eppig and the Mouse Genome Database Group+

The Jackson Laboratory, 600 Main Street, Bar Harbor, ME 04609, USA

Received October 1, 1998; Accepted October 6, 1998

ABSTRACT

The Mouse Genome Database (MGD) focuses on the integration of mapping, homology, polymorphism and molecular data about the laboratory mouse. Detailed descriptions of genes including their chromosomal location, gene function, disease associations, mutant phenotypes, molecular polymorphisms and links to representative sequences including ESTs are integrated within MGD. The association of information from experiment to gene to genome requires careful coordination and implementation of standardized vocabularies, unique nomenclature constructions, and detailed information derived from multiple sources. This information is linked to other public databases that focus on additional information such as expression patterns, sequences, bibliographic details and large mapping panel data. Scientists participate in the curation of MGD data by generating the Chromosome Committee Reports, consulting on gene family nomenclature revisions, and providing descriptions of mouse strain characteristics and of new mutant phenotypes. MGD is accessible at http://www.informatics.jax.org

INTRODUCTION

MGD continues to provide a public, community database resource of the laboratory mouse. This effort has four components: data acquisition, data integration, data querying and data visualization. The data represented in MGD can be viewed in several ways.

(i) As a mapping resource, the information about the mouse genome is represented in various maps built from experimental data submitted and integrated into MGD. These maps can be constructed with parameters defined by the user; some are consensus maps, and others are constructed by software algorithms from raw data. In addition, several types of comparative map displays can be constructed based on published mammalian homology data curated by MGD editors.

(ii) As a repository of experimental data, MGD records elemental information about mapping experiments, haplotypes recorded, probe details, and other information in support of map displays and for researchers to consult to evaluate the confidence for individual gene placement.

(iii) As a record of mouse genes, gene reports provide a robust set of information about a gene, its location, sequence, alleles, its homologs, gene products and function, and supporting references. These records are updated nightly resulting in timely summaries of the knowledge about an individual gene.

(iv) As a gene and allele nomenclature resource, MGD is queried by journals, scientists, and other database providers as the source of approved nomenclature symbols for mouse genes. The MGD nomenclature committee works extensively with nomenclature groups for human and other model organisms and with scientific committees working on providing a unique set of gene symbols and names for specific gene families.

(v) As an information source about the laboratory mouse in general, MGD provides information about characteristics and history of mouse strains, links to other mouse sequence and mapping resources, list-serve support, and documentation.

(vi) As a community resource, MGD personnel work with scientists to accept electronically submitted data sets and to represent these data to the scientific community in both their elemental form and as part of the integrated summary information available about the mouse genome and genes. MGD represents a consensus map position for genes and loci that is provided by community Chromosomal Committees and that is updated regularly.

CURRENT STATUS AND GENERAL ENHANCEMENTS

The numbers below were gathered from database statistics as of September 22, 1998. As this paper is an update, reporting new data and representations in MGD, readers are referred to previous publications for in depth coverage of MGD structural organization and overall coverage (1–3).

Genes and genetic markers

The number of genes and genetic markers detailed within MGD continues to increase rising to over 23 764 this year. This represents 20 660 mapped loci including over 6392 mapped

*To whom correspondence should be addressed. Tel: +1 207 288 6248; Fax: +1 207 288 6131; Email: jblake@informatics.jax.org

+The Mouse Genome Database Group: R. Baldarelli, J. Beal, R. Blackburn, D. Bradt, N. Butler, G. Colby, L. Corbani, J. Corrigan, C. Donnelly, J. Gilbert, L. Glass, P. Grant, M. Lennon-Pierce, L. Maltais, M. May, J. Merriam, J. Ormsby, S. Ramachandran, D. Reed, S. Rockwood and P. Trepanier.

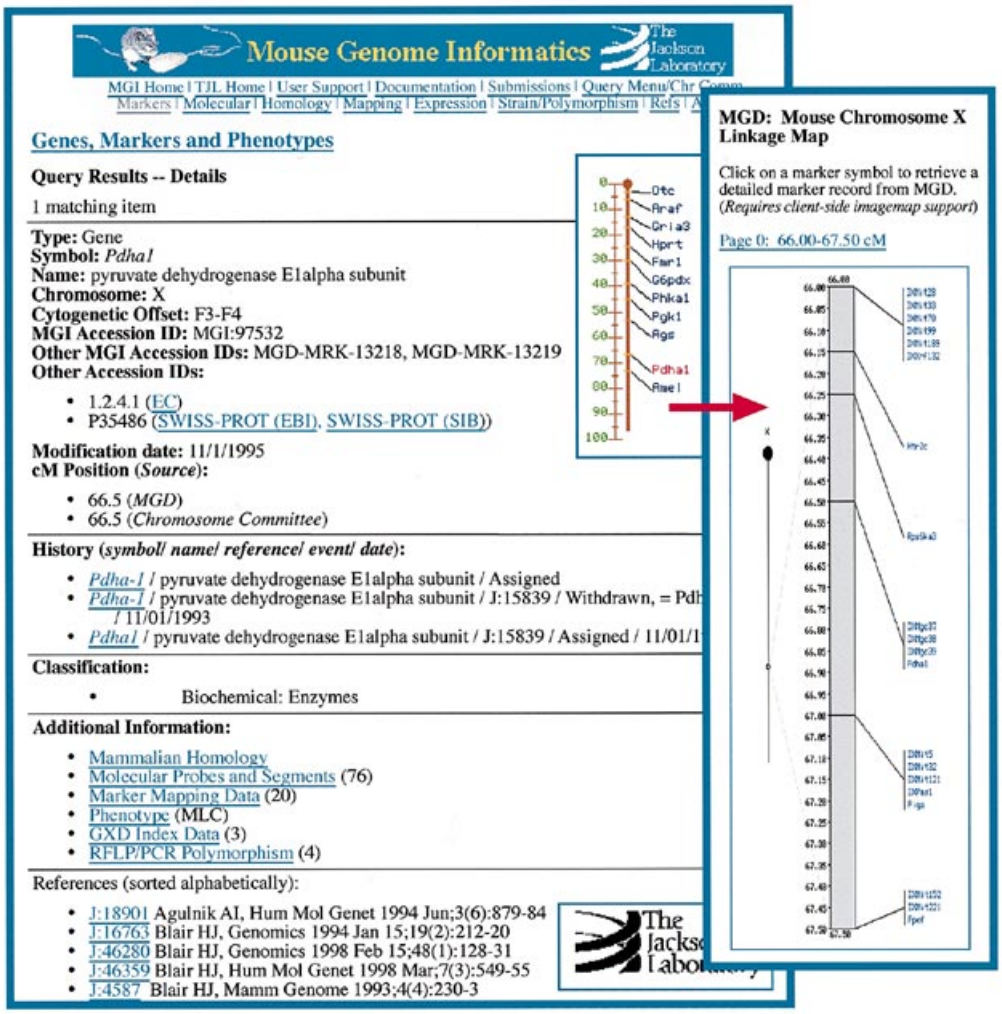

Figure 1. The enhanced linkage map display from the marker detail record allows retrieval of the web map display of the chromosomal region within 1 cM of the selected marker. All the markers in the region are displayed, including anonymous DNA segments. Each marker is linked to a marker detail record. This allows exploration of the markers that immediately surround a gene of interest without having to initiate a new search query.

genes. Genetic markers in MGD include genes, chromosomal aberrations, QTLs, anonymous DNA segments, and phenotypically defined Mendelian traits. Each marker detail contains a mini-map that shows the marker in association with other standard anchor loci on the chromosome. This display can be clicked to expand to a web map display of the chromosomal region and associated markers in the 1 cM region around the selected marker (Fig. 1). Over 3000 genes have been matched with their human ortholog and over 1170 with their rat ortholog, each with supporting documentation and evidence assertions (note that within MGD over 80 mammalian species have documented orthologs with some mouse genes).

Homology and comparative genomics

Mammalian homology data can be searched by species, gene symbol, name or map position. A nightly database report includes listings of all mouse/human homologies and mouse/rat homologies sorted by chromosome or gene symbol or by various other criteria. Comparative maps can be built using the linkage map building tools. The Oxford Grid display now includes a display option to choose between the genomes of two selected species. The Oxford Grid display also provides links to graphical comparative maps showing either a complete map or a region of a mouse chromosome with homologies from the selected comparison species. Homology between two selected species can be viewed as an Oxford Grid display that provides a genome-wide synopsis of homology relationships. Graphical comparative maps provide a detailed chromosomal view of conserved segments between mouse and other mammalian species.

Nomenclature

Gene nomenclature issues continue to play a tremendous role in the data curation process. A major component of the integration of genetic and genomic data in MGD depends on the curation of a unique set of gene symbols and names for the laboratory mouse. The MGD Nomenclature Committee works under the guidelines set by the International Committee on Standardized Genetic Nomenclature for Mice. Several journals now require review of

gene nomenclature as part of the manuscript review process. The MGD nomenclature coordinator works extensively with researchers to clarify and resolve gene nomenclature issues before publication. Scientists can obtain new gene symbols rapidly through the use of the nomen electronic submission form (http://www.informatics.jax.org/nomen/). Gene family information is often consolidated and posted on the Web by interested groups of scientists. MGD links to these Web pages and works closely with the scientists to resolve outstanding nomenclature issues. The curation of new mouse gene designations is often accomplished through coordination with the HUGO human gene nomenclature effort.

Mapping status and enhancements

Over 20 000 genetic markers are presented in the mapping representations of the mouse genome at MGD and the integration of many kinds of experimental mapping data continues to be a priority for MGD curators. Of the 9255 genetic markers identified as coding genes, 6392 have associated mapping data and are represented on the mouse genetic map. Among the mapping data sets available are 12 DNA Mapping Panels and composite sets of Recombinant Inbred (RI) Strain Distribution Patterns and Recombinant Congenic (RC) Strain Distribution Patterns. Graphical map displays are available for linkage, cytogenetic, and physical maps. Genetic maps can be constructed with user defined parameters including choice of data set and type of gene or marker (e.g. only genes involved in hearing processes). Maps can be built with cross-referencing to homologous genes of human or other mammalian species.

Revised 'Build a Linkage Map' form. The query form used to construct a linkage map has been reorganized and simplified so that a researcher can more easily generate a linkage map using information retrieved from the database. The view of the chromosome is determined by the criteria used to search the database. One can specify chromosome, a region of chromosome, include markers from selected phenotypic classes, include DNA segments or syntenic markers, or add your own markers. A map showing homologies with mouse can be generated by selecting a mammalian species for comparison. A new option allows one to choose to display only those mouse markers that have homologies in the selected species.

AXB/BXA mapping data sets available via FTP. Special purpose data sets were produced by Dr Beverly Paigen's laboratory at The Jackson Laboratory that increase known genotype information on the AXB and BXA RI mapping data sets. This led to development of a data set designed for QTL typing. In MGI 2.0, these data files are available in EXCEL format and can be downloaded from our ftp server. This information is also incorporated into MGD's composite RI data sets for AXB and BXA and is represented in the RI mapping experimental data records.

Community electronic data entry

Increasingly, MGD relies on direct submission of data sets (http://www.informatics.jax.org/doc/submit.html). This year, the guidelines for data submissions have been updated to facilitate Email submission of all kinds of MGD data. The electronic data submission process for mapping and phenotype data has been updated. A new feature is the development of procedures for the submission of mouse mutant phenotype descriptions for characterized remutations of known genes. User Support is available to help with data entry procedures as necessary. MGI accession numbers are assigned prior to publication so that they can be reported in the publication and link a publication to an electronically submitted dataset.

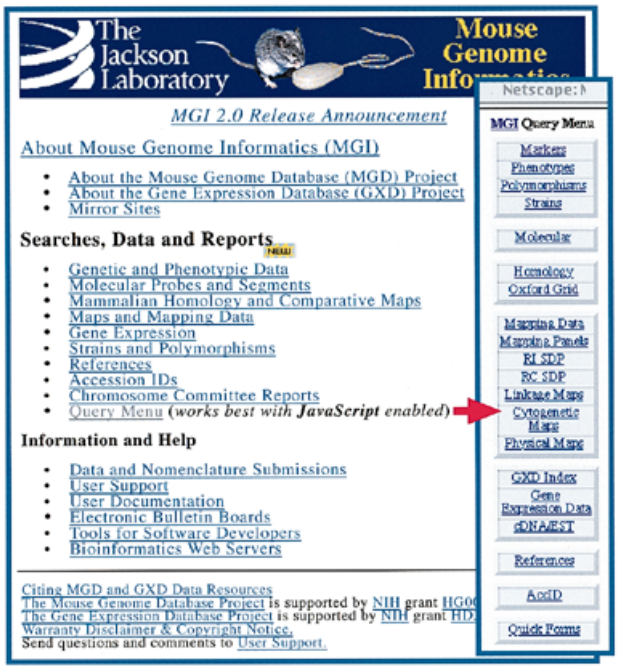

Figure 2. The new MGI home page provides access to the MGD and GXD through a set of menus organized with respect to biological information. A new pop-up Query Menu provides direct links to query forms.

SIGNIFICANT RECENT ENHANCEMENTS AND ADDITIONS

The Mouse Genome Informatics web site

From 1994 to 1998, the Mouse Genome Database was the sole data resource available from our web site. In 1998, the Mouse Genome Informatics (MGI) web site was developed to provide access to the Gene Expression Database (GXD) and to consolidate access to MGI data resources. The MGI web site provides integrated access to various information resources on the genetics and biology of the laboratory mouse, including the MGD, GXD (4), the Mouse Tumor Database (MTB) (5) and the Encyclopedia of the Mouse Genome. The new MGI home page (Fig. 2) provides links to high-level descriptive documentation, searches and reports, help, and other information. MGD is updated on a daily basis by the data curators. Significant database or software changes result in a version release. Over the course of the last year, MGI has released versions 1.0 and 2.0. Each release incorporated significant enhancements to the database structure and public accessibility.

Controlled vocabularies

Standardization of terms and vocabularies within MGD and GXD databases are being used to facilitate data entry and searching. We

continue to work intensively to provide standardization of gene and marker symbols and names. In addition, during the last year, we have normalized strain names and are providing continuous curation of strain and allele names in the database. Anatomical terms continue to be standardized. Source information (libraries) for ESTs includes standardized terms for age and sex. The classification of genes by gene function and biological process is underway.

Phenotype descriptions

Many genes have associated with them a gene report that summarizes current knowledge of the gene, its expression, product function, relationship to other genes and use of the gene as a model for human diseases. These gene reports are an outgrowth of the former Mouse Locus Catalog (MLC). During the last year attention was focused on the association of mouse genes with human phenotypes through the incorporation of OMIM (Online Mendelian Inheritance in Man) numbers as appropriate.

IMPLEMENTATION

MGD is implemented in the Sybase relational database system, version 11.0.3. The database consists of ~130 tables in which information about each type of entity (e.g., a mapping experiment) is cross linked to related information (e.g., the markers used) through unique identifiers. A large (and growing) set of stored procedures and triggers ensures continued integrity of the data in the face of updates. There are three major interfaces to MGD: editorial, WWW and SQL. The editorial interface is used by the MGI staff for entering and modifying database records. It is implemented with the TeleUSE user interface management system, a general-purpose tool for building X/Motif interfaces. The WWW interface comprises a set of static HTML forms and other supporting documents, and a large set of CGI (Common Gateway Interface) scripts that mediate the user's interaction with the database. Essentially, these scripts convert a completed query form or a selected hypertext link into an SQL query, execute the query, format the results as HTML, and return the result to the user. These scripts are written in Python, an object-oriented, interpreted language, available free on the Internet (http://www.python.org). Finally, users may access MGD directly via SQL. Users wishing an SQL account may contact MGD User Support.

ADDRESSES AND USER SUPPORT

URLs and mirror sites

MGD can be accessed at The Jackson Laboratory at http://www.informatics.jax.org . There are now five mirror sites around the world to provide users with faster local access to MGD:

UK: http://mgd.hgmp.mrc.ac.uk/
Japan: http://mgd.niai.affrc.go.jp/
France: http://www.pasteur.fr/Bio/MGD/
Australia: http://mgd.wehi.edu.au:8080/
Israel: http://bioinfo.weizmann.ac.il:3455/

Additional mirror sites are being considered. Mirror sites have the option of downloading FTP update files on a nightly basis. Most sites update less often, but on a regular basis.

Community outreach and user support

MGD provides extensive user support through on-line documentation and easy Email or phone access to user support staff. MGD staff attended over 35 meetings and symposia within the last year where posters, talks and demonstrations of the database were presented.

User Support WWW access: http://www.informatics.jax.org/doc/support.html
Email: mgi-help@informatics.jax.org
Telephone: 1 207 288 6445
FAX: 1 207 288 6132

MGI maintains an electronic bulletin board to promote communication among researchers on a variety of topics related to mouse genetic research and the MGI database resource. MGI-LIST has >1100 subscribers. Subscriptions can be obtained through the Web at http://www.informatics.jax.org/doc/lists.html

CITATION OF MGD

The following citation format is suggested when referring to specific datasets within MGD: Mouse Genome Database (MGD), Mouse Genome Informatics, The Jackson Laboratory, Bar Harbor, Maine (URL: http://www.informatics.jax.org). [Type in date (month, year) when you retrieved the data cited.]

To reference the database itself, please cite this article as well as others found at: http://www.informatics.jax.org/doc/citation.html

ACKNOWLEDGEMENT

The Mouse Genome Database and the Mouse Genome Informatics Project are supported by NIH grant HG00330.

REFERENCES

1 Blake,J.A., Richardson,J.E., Davisson,M.T., Eppig,J.T. and the Mouse Genome Informatics Group (1997) *Nucleic Acids Res*., **25**, 85–91.
2 Blake,J.A., Eppig,J.T., Richardson,J.E., Davisson,M.T. and the Mouse Genome Informatics Group (1997) *Nucleic Acids Res*., **26**, 130–137.
3 Eppig,J.T., Blake,J.A., Davisson,M.T. and Richardson,J.E. (1998) In *METHODS: A Companion to Methods Enzymol*., **14**, 179–190.
4 Ringwald,M., Mangan,M.E., Eppig,J.T., Kadin,J.A., Richardson,J.E. and the Gene Expression Database Group. (1998) *Nucleic Acids Res*., **27**, 106–112
5 Bult,C.J., Krupke,D.M. and Eppig,J.T. (1998) *Nucleic Acids Res*., **27**, 99–105.