MUGEN mouse database; Animal models of human immunological diseases

V. Aidinis*, C. Chandras, M. Manoloukos, A. Thanassopoulou, K. Kranidioti,
M. Armaka, E. Douni, D. L. Kontoyiannis, M. Zouberakis, G. Kollias[†] and the Mugen NoE
consortium[‡]

B.S.R.C. Alexander Fleming, 34 Fleming Street, 16672, Vari, Greece

ABSTRACT

The MUGEN mouse database (MMdb) (www.mugen-noe.org/database/) is a database of murine models of immune processes and immunological diseases. Its aim is to share and publicize information on mouse strain characteristics and availability from participating institutions. MMdb's basic classification of models is based on three major research application categories: Models of Human Disease, Models of Immune Processes and Transgenic Tools. Data on mutant strains includes detailed information on affected gene(s), mutant allele(s) and genetic background (DNA origin, gene targeted, host and backcross strain background). Each gene/transgene index also includes IDs and direct links to Ensembl, ArrayExpress, EURExpress and NCBI's Entrez Gene database. Phenotypic description is standardized and hierarchically structured, based on MGI's mammalian phenotypic ontology terms. Availability (e.g. live mice, cryopreserved embryos, sperm and ES cells) is clearly indicated, along with handling and genotyping details (in the form of documents or hyperlinks) and all relevant contact information (including EMMA and Jax/IMSR hyperlinks where available). MMdb's design offers a user-friendly query interface and provides instant access to the list of mutant strains and genes. Database access is free of charge and there are no registration requirements for data querying.

INTRODUCTION

The laboratory mouse serves as a premier animal model in studying the complex mechanisms involved in human disease. The genetic similarity of mice and humans has allowed the generation of a large reservoir of potential mouse models of human disease through the use of recent technological advances like gene cloning and transgenic technologies. Immunological diseases, in particular, encompass a wide variety of disorders of the immune system, including multiple sclerosis, systemic lupus, rheumatoid arthritis and cancer, all of which share basic mechanisms of initiation and progression. To better understand the complex mechanisms regulating the immune system and its pathological processes, it is important to systemically study animal models of immune diseases and processes through the application of functional genomic platforms, such as transgenesis, targeted random mutagenesis, expression profiling and bioinformatics. In this context, the MUGEN network of Excellence, a consortium of 21 leading research institutes and universities (Mugen NoE Consortium; www.mugen-noe.org/), developed the Mugen Mouse Database (MMdb; www.mugen-noe.org/database/), a virtual online mutant mouse strain repository, which currently holds all the mutant mouse models developed within the consortium. Its main goal is to publicize existing mouse models of immune diseases and processes, and to disseminate details of their particular characteristics. Ultimately, MMdb aims to use its particular expertise to become a database on an international scale. MMdb is hosted at the B.S.R.C. Alexander Fleming Institute (www.fleming.gr/)

*To whom correspondence should be addressed. Tel: +30 210 9654382; Fax: +30 210 9656563; Email: v.aidinis@fleming.gr
†Correspondence may also be addressed to George Kollias. g.kollias@fleming.gr

DATABASE DESIGN AND IMPLEMENTATION

The MMdb is the front end of a relational PostgreSQL (version 8.1.3) database. Currently the actual schema consists of 124 tables and is completely stripped down to its 5th normal form (5NF). The simplified relations between database entities are one of MMdb's biggest advantages. Adaptation of mouse related data structures is thus easier to implement, and complex queries can be readily handled by both the developer and the RDBMS.

MMdb's front and middle ends are entirely built on J2EE and have a greater degree of complexity. At the moment the MMdb consists of 1037 java source files, 560 JSP files, 268 html files and 61 xml files; this results in approximately 142 000 source lines of code (SLOC). Architecturally the application is divided in three layers: (i) The EJB layer, (ii) The Session layer and (iii) The Web layer.

The first layer provides the application with an object oriented API to the database. Significant database entities are represented with a corresponding EJB, which are then handled by the Session layer. All relational and combinational methods are executed by this second layer. The results are transferred to the Web layer, which handles the representation of data, implementing static and dynamic web pages.

The MMdb is currently deployed on Sun's System Application Server 8.2.

DATABASE AVAILABILITY

MMdb contents are freely accessible to all visitors and are not password protected. There is no need for visitors to login upon entering MMdb; on the contrary, they may browse the available data by clicking on the available links (i.e. Mugen mice, Gene, Model of Human Disease, Model of Immune Processes, Transgenic Tool) on the homepage or via the search option, available in MMdb at all times.

DATA ENTRY AND CURATION

Any type of data that MMdb maintains can be submitted in electronic form. Currently, data submission and modification is only possible for registered users as described below: (i) *MMdb administrators* have full access rights to all of the functionalities and contents of the database. (ii) *MUGEN registered users* have rights to data entry and modification only for data entered by themselves, and are unable to alter data entered by other Mugen registered users under any circumstances.

The MMdb curation team surveys journals and other online resources, collects all the data and processes the submission of each mutant mouse individually prior to its deposition into the database. Submissions are checked for accuracy and completeness of biological information, including genetic, strain, allelic and mutational features. Once curation of a particular mouse is finalized the MMdb development team, in close collaboration with the responsible researcher, works towards bringing the data into standardized formats, resolving issues pertinent to

nomenclature and referential integrity, in order to ensure data accuracy and correctness. Description of this curation method is also described on MMdb's 'about page'. Mutant mouse data are currently under regular curation and are continuously revised, in order to maintain a constantly updated version of MMdb for the user to refer to. The date that the information for the respective mouse was last updated is also available for the user.

DATA CATEGORIZATION

Detailed information is provided for each mutant strain. These are categorized according to the context of the details offered. The first section presents in detail general information on the particular mutant strain: MUGEN ID, the common name in general use for the particular mutant mouse, contact person responsible for the mutant and institution that he/she is affiliated with (Figure 1). Each mutant strain is assigned a research application type denoting if the particular model has been created to serve as a model of 'human disease', 'immune processes' (i.e. for examining signalling pathways and cellular components involved in immune responses) or a 'transgenic tool' (i.e. transgenic mouse lines that provide a powerful tool for generating conditionally mutagenized mouse lines). Finally, relevant additional comments that are entered by the contact person for the respective mutant mouse are also presented. Availability details allow the user to view the repository where the particular mutant strain is housed, as well as its genetic background, strain type and state information (Figure 1).

In addition to providing basic information for the mutant strains, MMdb also offers more detailed features like the DNA origin, targeted and host background, as well as the backcrossing strain and number of backcrosses, thus covering features of the genetic background of the mutant strain (Figure 1). Supplementary strain information, such as the exact strain designation, is provided, together with the JAX or EMMA IDs (where available) in conjunction with the direct links to the respective databases. Furthermore, allelic characteristics are also offered in a separate tab, where the standard allele symbol and name are presented together with the corresponding MGI ID (where available) and the direct link to the MGI database. Moreover, basic information for the gene/transgene, like the gene name, symbol and related chromosome are presented to the user, together with a list of Mugen mice available in MMdb which have a mutation relevant to the particular gene (Figure 2). In addition, each gene/transgene index also includes IDs and direct links to Ensembl [(1) www.ebi.ac.uk/ensembl/], ArrayExpress [(2) www.ebi.ac.uk/arrayexpress/], EURExpress II (www.eurexpress.org/ee/) and NCBI's Entrez Gene [(3) www.ncbi.nlm.nih.gov/sites/entrez?db = gene] databases. Overall, gene, allelic and mutational nomenclature is assigned according to rules and guidelines for mouse genes and strains given by MGI [(4) www.informatics.jax.org/mgihome/nomen/strains.shtml] and based on Gene

Figure 1. Available information and data categorization on a sample mouse model of MMdb. Representative view of a mutant mouse 'Tnfalpha$^{-/-}$' from MMdb, demonstrating all the individual data categorizations and information that are recorded in MMdb (general information, availability, genetic background, strain information, allele and mutations, genes affected, handling and genotyping instructions, phenotype(s) references, author's comments and user's comments).

gene

help & info

Name: tumor necrosis factor
Symbol: Tnf
Synonym/Old Name: tumor necrosis factor alpha
Chromosome: 17
MGI ID: MGI:104798
Entrez ID: 21926
ENSEMBL ID: ENSMUSG00000024401
ArrayExpress: expression profiles
Eurexpress: embryonic expression pattern
Comment
not provided yet

Related MUGEN Mice

TNFalpha$^{-/-}$

Tnf^{tm}

TNF^{deltaAREneo}

TNF^{deltaARE}

Related Alleles

Tnf^{tm1Gkl}

Tnf^{tm3Gkl}

Tnf$^{tm2.1Gkl}$

Tnf^{tm2Gkl}

Related MP Terms

adipose tissue phenotype>abnormal adipose tissue physiology

growth/size phenotype>abnormal postnatal growth/weight/body size>abnormal body size>abnormal body weight>decreased body weight>weight loss

growth/size phenotype>abnormal postnatal growth/weight/body size>abnormal body size>decreased body size>decreased body weight

digestive/alimentary phenotype>abnormal digestive system physiology>intestinal inflammation>small intestinal inflammation

immune system phenotype>abnormal immune system morphology>abnormal immune system cell morphology>abnormal white blood cell number>increased white blood cell number

immune system phenotype>abnormal immune system physiology>abnormal humoral immune response>abnormal immunoglobulin concentration>decreased immunoglobulin concentration>decreased IgG level

immune system phenotype>abnormal immune system physiology>abnormal inflammatory response>increased inflammatory response>intestinal inflammation>small intestinal inflammation

immune system phenotype>abnormal immune system physiology>abnormal inflammatory response>increased inflammatory response>joint inflammation>chronic joint inflammation

immune system phenotype>abnormal immune system physiology>abnormal inflammatory response>increased inflammatory response>joint inflammation>arthritis

immune system phenotype>abnormal immune system physiology>abnormal inflammatory response>decreased inflammatory response

immune system phenotype>abnormal immune system physiology>abnormal inflammatory response>abnormal acute inflammation

immune system phenotype>abnormal immune system physiology>abnormal inflammatory response>abnormal inflammatory mediator physiology>abnormal cytokine physiology>abnormal TNF physiology

immune system phenotype>abnormal immune system physiology>abnormal adaptive immunity

immune system phenotype>abnormal immune system physiology>abnormal immune serum protein physiology>abnormal inflammatory mediator physiology>abnormal cytokine physiology>abnormal TNF physiology

immune system phenotype>abnormal immune system physiology>immune tolerance>abnormal self tolerance>autoimmune response>altered susceptibility to autoimmune disorder>decreased susceptibility to autoimmune disorder>decreased susceptibility to experimental autoimmune encephalomyelitis

immune system phenotype>abnormal immune system physiology>abnormal response to infection>altered susceptibility to infection>increased susceptibility to infection>increased susceptibility to bacterial infection

immune system phenotype>abnormal immune system physiology>abnormal response to infection>altered susceptibility to infection>decreased susceptibility to infection>decreased susceptibility to bacterial infection

skeleton phenotype>abnormal skeleton morphology>abnormal joint morphology>joint inflammation>chronic joint inflammation

skeleton phenotype>abnormal skeleton morphology>abnormal joint morphology>joint inflammation>arthritis

skeleton phenotype>abnormal skeleton morphology>abnormal joint morphology>joint swelling

hematopoietic system phenotype>abnormal hematopoietic system morphology/development>abnormal thymus morphology>thymus atrophy

back

Figure 2. Available information on a sample gene of MMdb. Representative view of a gene 'tumor necrosis factor' from MMdb, demonstrating basic information for the gene/transgene together with a list of Mugen mice, alleles and MP terms available in MMdb, which are relevant to the particular gene. Each gene/transgene index also includes IDs and direct links to Ensembl, ArrayExpress, EURExpress II and NCBI's Entrez Gene database.

Ontology (GO) [(5) www.geneontology.org], Mammalian Phenotype (MP) terminology (6–8) and international nomenclature standards.

Handling and genotyping information, in the form of documents or hyperlinks, are also provided in MMdb. Recognizing that phenotype data are by nature complex and usually incomplete, the MP Ontology has been developed to describe the richness and complexity of phenotypes more precisely and to support the goal of comparative phenotyping and building new animal models. MP terms are used as a structured vocabulary and a tool for annotating, analyzing and comparing phenotypic information. Where MP terminology is available for Mugen mice the phenotypic descriptions are standardized and hierarchically structured according to MGI's MP terms, which are updated weekly by MMdb's informatics team (Figure 1).

Other features provided by MMdb for each mutant mouse strain include relevant references in the form of hyperlinks or documents, and additional comments

provided by the author. Furthermore, MMdb offers a specific tab within each mutant strain, where any visitor can submit additional comments. This feature serves as a forum for discussion between the visitors and intends to promote exchange of useful information between them.

The MMdb currently holds (142) listed mutant mice and (165) related genes.

QUERYING THE DATABASE

MMdb provides a browsing/filtering interface to the underlying data, allowing formulation of queries, in order to narrow down the list of potential results and to coordinate screening of certain mutant strains. Automated drop-down menus that include research applications (model of human disease, model of immune processes, etc.), genes (tumor necrosis factor, CD19 antigen, etc.), institutions, researchers/contact persons or mutation types (insertion, duplication, deletion, targeted, etc.), assigned MP terms (abnormal T cell physiology, abnormal spleen morphology and other already assigned MP terms) or

even a combination of the above selections allow further refinement of the mutant strain list according to special characteristics defined by the user (Figure 3A).

Another way in which stored data can be queried is via the function formulating *ad hoc (string)* queries. At all times, the user can enter the desired query word(s) into the available text box and seek for the particular word(s) at all levels of the database. In all cases, a chart is automatically generated with the corresponding mutant strains and genes, which in turn can be accessed by simply clicking on the line name of the particular mutant strain or gene (Figure 3B).

A detailed help section is also available in MMdb. In-depth directions on how to view different parts of MMdb are provided. These are categorized in three different sections, (i) View Mugen Mutant Mouse Details, (ii) View Gene Details and (iii) Search the Mugen Database, all of which are expandable by clicking on each of the available tabs. Under each of the offered options, additional, step by step information is presented, aiming to provide directions for the user to make the best use of MMdb.

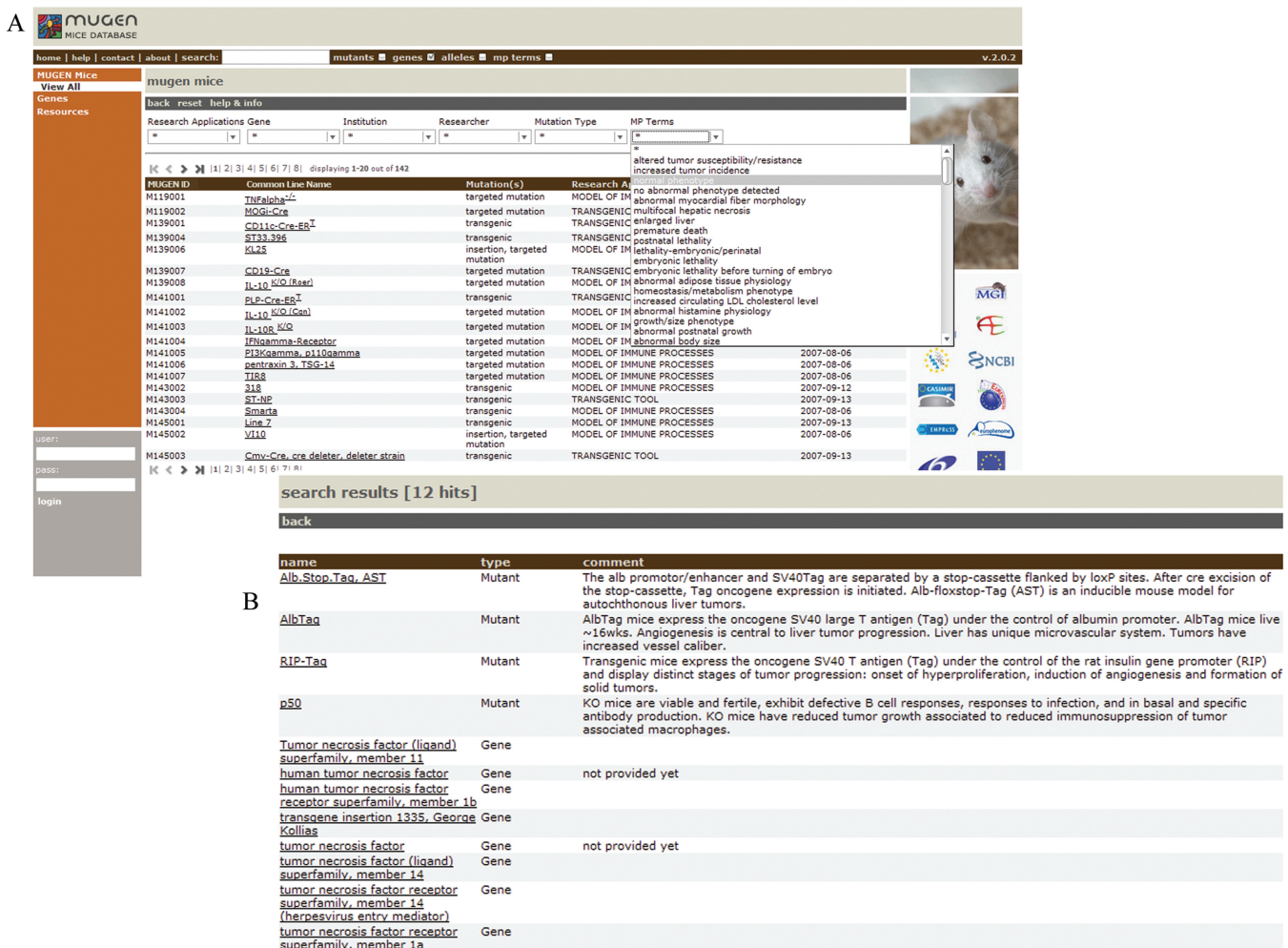**Figure 3.** Querying MMdb. (**A**) A screen shot demonstrating how database querying can be restricted via choosing from drop down menus (e.g. research application, gene, institution, researcher, mutation type dissemination level and MP term) and also via the function formulating *ad hoc* queries. (**B**) Results of an *ad hoc* query for 'tnf'. A list with 12 hits of all the available mutant mouse models and related genes is available.

OTHER DATABASE CHARACTERISTICS

User feedback is fundamental for improving overall database usability and data accuracy. As a result, we have created a specific tab for visitors to email the MMdb developing team (mmmdb@mugen-noe.org) with any comments, on either the website in general or a particular mutant mouse model. MMdb users' opinions are extremely useful to us, as we can improve our database, data presentation and update them according to the feedback obtained.

Furthermore, MMdb can serve as a reservoir of useful information on all online mouse-related resources and thus holds a well categorized and comprehensive list of links to external resources. The associated tab may be found on the left hand side of MMdb's page at all times and will immediately redirect the user to the list of available resources. Furthermore, MMdb maintains direct links to the home pages of IMSR [(9) www.informatics.jax.org/imsr/], MGI (www.informatics.jax.org/), Ensembl, ArrayExpress, EMMA (www.emmanet.org/), NCBI [(10) www.ncbi.nlm.nih.gov/sites/entrez], CASIMIR (www.casimir.org.uk/) and EURExpress, which may be found in the form of icons on the right hand side of MMdb and will transfer the user directly to the requested link.

CONCLUSION/DISCUSSION AND FUTURE PROSPECTS

MMdb is a comprehensive and user-friendly source of information for exploring genetic and phenotypic information on mutant mice of human immunological disease. Core data include affected gene, genetic background, strain, allelic, mutational and phenotypic information as well as details on the handling and genotyping of mice. Additional information, in the form of comments inserted by the author, are also provided together with links to related published papers. Data are integrated through a combination of expert human curation, and use a variety of controlled/structured vocabularies (ontologies), like the MP Ontology.

Although, MMdb is freely accessible to all individuals, data submission is restricted to registered users for the moment. However, MMdb aims to expand significantly by increasing the number of end-users and data contributors from leading institutes and universities. Ultimately, MMdb's goal is to become an international level reference source for information on models of immunological disease and immune processes.

MMdb is constantly being updated at the curation level, to enrich the existing data collection, but also to upgrade the MMdb software. First and foremost, taking into account the necessity for database interoperability and combinational data from different resources, we intend to develop web services. The improved ability to combine relevant information from different online resources will provide additional value to each of the individual interoperable databases, but most importantly to the whole community using mouse as a model organism. MMdb, through an ongoing close collaboration with Ensembl, aims to enhance their BioMart software and connect it directly to MMdb, tracing our mouse strains based on their allele, background and affected gene. Our ultimate goal is to provide as much information as possible from the mouse genome through to the mouse phenome.

The MP ontology and annotation schema has been designed over the past couple of years to support robust phenotypic annotations and querying capabilities for mouse phenotype data. MMdb, staying up to date with phenotypic descriptions, is currently developing an MP term look up service, thus enabling its users to detect MP terms and their entire relative paths via simple keyword(s) search, and add desired MP ontologies to the terminology collection of specified mutant strains. Our goal is to eventually make phenotypic assignment as simple as possible, even for users who are not familiar with the constantly growing coverage of MGI's MP Ontology.

Database content management and data accuracy are some of the key elements for successful database projects. Moreover, sustainability and effectiveness of the database and informatics infrastructure, as measured by the willingness of scientists to use them and deposit their data in such databases for the scientific benefit of the whole community, is of considerable value for a successful database. In order to provide incentives to potential contributors and researchers to become registered users and submit their mutant mouse strains to MMdb, our aim is to continue acting as a specialized database and to provide a high standard repository, exclusively for mouse models of immune processes and immunological disease.

More importantly, MMdb's ultimate goal is to create a mouse-centric international forum on modelling of immunological diseases and promote correlation of various genotypic and phenotypic characteristics, which would pave the way to systems biology of the mouse.

REFERENCES

1. Hubbard,T.J.P., Aken,B.L., Beal,K., Ballester,B., Caccamo,M., Chen,Y., Clarke,L., Coates,G., Cunningham,F. *et al.* (2007) Ensembl 2007. *Nucleic Acids Res.*, **35**, D610–D617.
2. Parkinson,H., Kapushesky,M., Shojatalab,M., Abeygunawardena,N., Coulson,R., Farne,A., Holloway,E., Kolesnykov,N., Lilja,P. *et al.* (2007) ArrayExpress–a public database of microarray experiments and gene expression profiles. *Nucleic Acids Res.*, **35**, D747–D750.

3. Maglott,D., Ostell,J., Pruitt,K.D. and Tatusova,T. (2007) Entrez Gene: gene-centered information at NCBI. *Nucleic Acids Res.*, **35**, D26–D31.

4. Eppig,J.T., Blake,J.A., Bult,C.J., Richardson,J.E., Kadin,J.A. and Ringwald,M. (2007a) Mouse genome informatics (MGI) resources for pathology and toxicology. *Toxicol. Pathol.*, **35**, 456–457.

5. Hill,D.H., Davis,A.P., Richardson,J.E., Corradi,J.P., Ringwald,M., Eppig,J.T. and Blake,J.A. (2001) Strategies for biological annotation of mammalian systems: implementing gene ontologies in mouse genome informatics. *Genomics*, **74**, 121–128.

6. Gkoutos,G.V., Green,E.C., Mallon,A.M., Hancock,J.M. and Davidson,D. (2005) Using ontologies to describe mouse phenotypes. *Genome Biol.*, **6**, R8.

7. Smith,C.L., Goldsmith,C.A.W. and Eppig,J.T. (2005) The mammalian phenotype ontology as a tool for annotating, analysing and comparing phenotypic information. *Genome Biol.*, **6**, R7.

8. Eppig,J.T., Blake,J.A., Bult,C.J., Kadin,J.A., Richardson,J.E. and Mouse Genome Database Group (2007b) The mouse genome database (MGD): new features facilitating a model system. *Nucleic Acids Res.*, **35**, D630–D637.

9. Strivens,M. and Eppig,J.T. (2004) Visualizing the laboratory mouse: capturing phenotype information. *Genetica*, **122**, 89–97.

10. Wheeler,D.L., Barrett,T., Benson,D.A., Bryant,S.H., Canese,K., Chetvernin,V., Church,D.M., DiCuccio,M., Edgar,R. *et al.* (2007) Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.*, **35**, D5–D12.