

IMSR File Format

Author: Mark Airey

Last Modified: September 02, 2010 01:47

1 IMSR Provider Data-Files

Institutions that contribute to the IMSR will provide their data on a regular schedule as tab-delimited data files. The file will identify a list of strains or stocks that are held at and distributed from the provider's site, and the "states" in which these strains or stocks are held. The list of valid states supported by the IMSR can be found in Table 3. The definition of the columns that should be provided in the tab-delimited IMSR load files can be found in Table 1, below. A single record/ line of data in the file can be thought of as having two parts: the strain specific fields (strain id, strain designation, strain type and strain state, all required); and fields with genetic information associated with that strain (allele accession/symbol/name, gene accession/symbol/name and chromosome, all optional). Strain specific fields are white in the table below; genetic information is shaded. One or more sets of genetic information may pertain to a given strain. If a strain carries multiple mutant alleles and/or chromosomal aberrations, a separate line is required for each set of genetic data. The same strain information should be contained in each of these lines.

Table 1: New format of IMSR Data File

Column in data-file	Definition
Strain/Stock ID	The unique ID used by a site to represent a strain. (Required)
Strain/Stock Designation	Full strain name. Superscripted portions of the name are enclosed in < >. For more information on nomenclature* of strain names see the Mouse Nomenclature Guide at http://www.informatics.jax.org/mgihome/nomen/index.shtml (Required)
Strain type	Type of strain, comma-separated list. See Table 2 for the accepted list of strain types. (Required)
State	State in which the strain is held. Like strain type, if a strain is held in multiple states at a given site, it should be included as a comma-separated list. See table 3 for the strain state abbreviations. (Required)
URL	The url parameter value, or full url used to create a link to strain detail web page at the repository site. (Required)
MGI Allele Accession ID	This is the MGI Accession ID for the Allele. Use of this field should make providing allele symbol and name unnecessary. (Optional)
Allele Symbol	Allele associated with this gene and strain. If a strain carries multiple alleles, a separate line is required for each.

	Please verify nomenclature* by searching MGI.
Allele Name	Name of Allele.(Optional)
Mutation Type	For each mutant allele, specifies the type of mutation that it represents. See Table 4 for the mutation type abbreviations. Usually there will be only one mutation type per Allele. Should there be an exception to this rule, these should be submitted as a comma-separated list. (Optional)
Chromosome	Chromosome on which the Gene is located, or the Chromosome(s) involved if the Mutation Type is CH (chromosomal aberration), RB (Robersonian Translocation) or TL (reciprocal translocation). If more than one chromosome is involved, they should be listed on the same line, in numerical order, separated by a comma. Valid chromosomes are numbers 1-19 plus X, Y, and: UN (unknown), MT(mitochondrial), XY(pseudoautosomal). (Optional)
MGI Gene Accession ID	This is the MGI Accession ID for the Gene. Use of this field should make providing gene symbol and name unnecessary. (Optional)
Gene Symbol	Gene Symbol for mutation carried in this strain. If a strain carries multiple gene mutations, a separate line is required for each. Please verify nomenclature* by searching MGI. (Optional)
Gene Name	Name of Gene. (Optional)

*Additional nomenclature assistance is available by emailing strains@informatics.jax.org.

The following tables (Table 2, 3, and 4) contain the abbreviations that should be used in provider data-files for strain type, strain state and mutation type.

Table 1: Valid Strain Types

Abbreviation	Strain Type
CCO	closed colony
COI	coisogenic strain
CON	congenic strain
CSS	consomic or chromosome substitution strain
IS	inbred strain
MAH	major histocompatibility congenic
MIH	minor histocompatibility congenic
MSK	mutant stock
MSR	mutant strain
NON	noninbred stock

RC	recombinant congenic
RI	recombinant inbred
SEG	segregating inbred
UN	unclassified
WDS	wild-derived inbred strain

Definitions for terms in the tables below can be found in the Glossary in section 3.

Table 1: Strain States

Abbreviation	State
EM	cryopreserved embryos
ES	ES cell lines
LM	live mouse
OV	cryopreserved ovaries
SP	cryopreserved or freeze-dried sperm

Table 1: Mutation Types

Abbreviation	Mutation Type
CH	chromosomal aberration
CI	chemically induced mutation
DEL	deletion
DP	duplication
GT	gene trap
INS	insertion
INV	inversion
OTH	other
RAD	radiation induced mutation
REC	recombinase(cre/flip)
RB	Robertsonian translocation
SM	spontaneous mutation
TG	transgenic
TL	reciprocal translocation
TM	targeted mutation
TP	transposition

A single line in a data file represents a distinct strain/gene/allele combination. If a strain is associated with multiple genes and/or alleles, there will be multiple lines in the file for that

strain.

Example: For this strain example (C57BL/6J-Rag1^{tm1Mom} Ins2^{Akita}/J), the submission file would include 2 lines. This is because two mutant alleles are carried by the strain.

```
4369 C57BL/6J-Rag1<tmlMom> Ins2<Akita>/J SEG,MSR LM,EM MGI:1857241 Rag1<tmlMom> tar-
geted mutation 1, Peter Mombaerts TM 2 MGI:97848 Rag1 recombination activating gene 1
4369 C57BL/6J-Rag1<tmlMom> Ins2<Akita>/J SEG,MSR LM,EM MGI:1857572 Ins<Akita> Akita
SP 7 MGI:96573 Ins2 insulin II
```

Table 1: Additional Provider Information

Each contributing site will need to provide specific information prior to its participation.

1. “Facility” information regarding the site. Facility information is submitted via the online registration form <http://www.informatics.jax.org/imsr/IMSRRRegistration.jsp> . A complete definition of this information can be found below in the description of the “Facility” table in the IMSR database.
2. Identification of the field in the data-file that will be used to complete the URL that links to detail about a specific strain at their facility. In the case of JAX this is the “stock id” (see further explanation under 2.1 Facility Table).
3. Contributed files should be placed in the “imsr/incoming” directory under <ftp.informatics.jax.org>. **File names should be of the form: JAX_04_2004**, where the name includes the site’s abbreviation and the month and year of submission. You will be provided with an ftp account to upload your files to our ftp site. If submissions will be more frequent than monthly, the day should be added to the name as well. Updated data files should be submitted at least every 6 months, although we encourage more frequent updates.

1.1 The Facility Table

A record in this table represents a facility data to the IMSR.

baseURL	varchar(255)	not null
siteName	varchar(255)	not null
institution	varchar(125)	null
addr1	varchar(125)	not null
addr2	varchar(125)	null
addr3	varchar(125)	null
addr4	varchar(125)	null
postalCode	varchar(20)	null
country	varchar(20)	not null

Phone	varchar(25)	not null
Fax	varchar(25)	not null
EMail	varchar(50)	not null
IMSRContactName	varchar(100)	not null
IMSRPhone	varchar(25)	not null
IMSRFax	varchar(25)	not null
IMSREmail	varchar(50)	not null
abbrevName	varchar(5)	not null

The *baseURL* attribute, paired with a strain specific piece of data, forms the total URL which takes the user to a page at the submitting facility giving further strain information. The *baseURL* attribute holds the part of the URL common to all strains in a facility. For example, if a strain's total URL is <http://jaxmice.jax.org/jaxmice-cgi/jaxmicedb.cgi?objtype=pricedetail&stock=4125>, the *baseURL* attribute would hold <http://jaxmice.jax.org/jaxmice-cgi/jaxmicedb.cgi?objtype=pricedetail&stock=###STRAIN###>. The value `###STRAIN###` is intended as a place-holder, so that the site/strain specific url (column 5 of data file) value can be inserted into the base URL appropriately. In our example the site/strain value is 4125.

The *siteName* attribute identifies the facility holding the strain. (ie. JaxMice)

The *institution* attribute identifies a super-entity of the facility specified by *siteName*. For instance, if a hospital laboratory were submitting strains, the *siteName* would be the laboratory name; *institution* would hold the hospital name. (ie. The Jackson Laboratory)

The Phone, Fax and Email attributes will be used in the IMSR web interface as the contact information to obtain information about mice/strains at a given facility.

The *siteName*, *institution*, *addr1-4*, *postalCode*, *country* and *Email* attributes are displayed on the IMSR webpage containing facility contacts.

The *IMSRContactName*, *IMSRPhone*, *IMSRFax* and *IMSREmail* attributes identify the individual acting as a liaison from the facility to the IMSR administrator at the Jackson Laboratory. These items are NOT displayed on the IMSR webpage.

The *abbrevName* attribute holds a three to five letter abbreviation for a facility name. (ie. JAX, HAR or MMRRC)

2 Glossary

This is a copy of the IMSR Glossary that will be made available as part of the IMSR web interface.

Accession ID

A unique alphanumeric character string that is used to unambiguously identify a particular record in a database. For example, MGI:96677 is the accession ID for the kit oncogene (*Kit*) in the Mouse Genome Informatics database; Y00864 is the GenBank ID for a mouse *Kit* mRNA sequence.

Allele

One of the variant forms of a gene, differing from other forms in its nucleotide sequence.

Allele Name

A word or phrase that uniquely identifies an allele of a given gene. The allele name has an abbreviation, the allele symbol, added as a superscript to the gene symbol. For example, one of the alleles of the *Kit* gene is named "viable dominant spotting" with a symbol, *Kit*^{W-v}.

Allele Symbol

A unique abbreviation for the allele name. Allele symbols take the form of superscripts added to the gene symbol. For example, one of the alleles of the *Kit* gene is named "viable dominant spotting" with a symbol, *Kit*^{W-v}.

Autosome

Any chromosome that is not a sex chromosome; for mouse, Chromosomes 1-19

Chemically Induced Mutation (CI)

A mutation induced by treatment with a chemical mutagen, for example, ENU (ethyl nitrosourea) or chlorambucil.

Chromosome

A structural unit within a eukaryotic nucleus that carries genes. A chromosome consists of a long, continuous strand of DNA and associated proteins.

Chromosomal Aberration (CH)

A chromosomal rearrangement or a numerical chromosomal aberration.

Chromosomal Rearrangement

A kind of mutation in which there is a change to the normal arrangement of the genome into chromosomes. Usually, but not exclusively, this term applies to those changes that are visible cytogenetically. Classes of chromosome rearrangements include: Deletions, Duplications, Insertions, Inversions, Translocations, and Transpositions. See also these specific terms.

Closed colony (CCO)

A colony propagated by a breeding protocol that seeks to maintain as many alleles segregating as possible without introducing new alleles from outside the colony or losing alleles through inbreeding.

Coisogenic (COI)

A strain that differs from its progenitor inbred strain at only one locus. A coisogenic strain arises when a mutation occurs in an inbred strain.

Congenic (CON)

An inbred strain that contains a small genetic region (ideally a single gene) from another strain, but which is otherwise identical to the original inbred strain. Congenic strains are derived by backcrossing to a parental inbred strain for at least ten generations while selecting for retention of a specific locus from the donor strain. For example, CBA/Ca.C3-*KitW-x*/J is an inbred strain of CBA/Ca background onto which the *KitW-x* allele from C3H/HeJ has been placed.

Consomic (CSS)

An inbred strain that contains a single entire chromosome from another strain. Consomic (or Chromosome Substitution Strains) are derived by backcrossing to a parental inbred strain for at least ten generations while selecting for retention of a specific whole chromosome from the donor strain. For example, C57BL/6J-Chr 4A/J/NaJ is a consomic strain in which the Chromosome 4 from the A/J strain has been placed on a C57BL/6J background.

Cryopreservation

The process by which living cells, organs, or organisms are frozen and maintained with the intent to revive them into functional units in the future.

Cryopreserved embryos (EM): the frozen state of embryos, usually frozen at the 2- or 8-cell stage.

Cryopreserved ovaries (OV): the frozen state of ovaries (female reproductive organ containing the gametes).

Cryopreserved sperm (SP): the frozen state of sperm.

Deletion (DEL)

A type of mutation caused by loss of one or more nucleotides from a DNA segment. Deletions can be very large, encompassing many genes and megabases of DNA, to the point of producing a visible cytological abnormality in a chromosome. Small deletions within a gene can alter the reading frame, and thus the amino acid sequence of the encoded protein.

Duplication (DP) An additional copy of a DNA segment present in the genome. Duplications can lead to an increase in the number of genes carried on a chromosome and may or may not be cytologically visible.

Embryos, cryopreserved (EM). See cryopreservation.

Embryonic Stem (ES) Cell

Undifferentiated cells derived from blastocyst-stage embryos. ES cells can differentiate in culture to a number of different cell types. When incorporated into chimeric mouse embryos they are totipotent and can differentiate into any cell type in the mouse. If these cells become part of the germline, they can mature to gametes and be "rescued" as living mice.

Gene

The basic unit of heredity; a portion of DNA that (usually) codes for a protein product. The term "gene" also may be used to refer to a locus in the cytoplasmic or nuclear genome characterized by an altered phenotype or by an effect on an inserted reporter gene, such as a gene trap or enhancer trap.

Gene Name

A gene name is a word or phrase that uniquely identifies a gene. The gene name has an abbreviation that is the gene symbol.

Gene Symbol

A gene symbol is a unique abbreviation for the gene name.

Gene Trap (GT)

A type of DNA construct containing a reporter gene sequence downstream of a splice acceptor site that is capable of integrating into random chromosomal locations in mouse. Integration of the gene trap into an intron allows the expression of a new mRNA containing one or more upstream exons followed by the reporter gene. The reporter gene is therefore expressed in the same cells and developmental stages as the gene into which the gene trap has inserted.

Inbred Strain (IS)

An inbred strain is essentially homozygous at all loci and each member of the strain is genetically identical to other members (excepting sex differences). In mice, an inbred strain is produced from brother x sister matings for at least 20 sequential generations. C57BL/6J is a widely-used inbred strain of mouse.

Insertion (INS)

A type of mutation in which one or more nucleotides is inserted into a DNA sequence. Small insertions within a gene can alter the reading frame, and thus the amino acid sequence of the encoded protein.

Inversion (INV)

A type of mutation in which a length of DNA is broken in two positions and repaired in such a way that the medial segment is now present in reverse order. Inversions range in size from those large enough to be visible cytogenetically to those involving only a few base pairs.

Knock in

A type of targeted mutation in which an alteration in gene function other than a loss-of-function allele is produced.

Knock out

A type of targeted mutation in which a loss-of-function (often a null allele) is produced.

Major histocompatibility congenic (MAH)

A congenic strain in which the donor allele transferred to the host strain background is a major histocompatibility gene.

Minor histocompatibility congenic (MIH)

A congenic strain in which the donor allele transferred to the host strain background is a minor histocompatibility gene.

Mutant

A term applied to a gene or phenotype altered by mutation or to an individual carrying a mutation.

Mutant Strain (MSR)

An inbred strain carrying one or more phenotypic mutations. For example, the mutant strain C3H/ H3eSn-bc3JJ carries the bouncy 3J phenotypic mutation.

Mutant Stock (MSK)

A stock carrying one or more phenotypic mutations. The term "STOCK" is used for a mutant strain, incipient or inbred, derived from more than two progenitor strains or having genetic contribution from an unknown genetic background source and is considered a "mixed" inbred. For example, the mutant stock named STOCK Col2a1*sedc*/J carries the spondyloepiphyseal dysplasia congenita phenotypic mutation of the collagen 2a1 gene.

Mutation

Any permanent change in DNA, i.e., in its nucleotide sequence. Examples include chromosome rearrangements and point mutations. The term mutation also is used for the process through which genes undergo a structural change.

Noninbred Stock (NON)

A stock that is known to be noninbred, or intentionally kept noninbred.

Ovaries, cryopreserved (OV). See cyropreservation.

Phenotype

A description of the observable state of an individual with respect to some inherited characteristic. Often, individuals with different genotypes display the same phenotype.

Radiation Induced Mutation (RAD)

A mutation induced by treatment with irradiation, in mouse usually gamma ray or X ray.

Recombinant Congenic Strains (RC)

Strains formed by crossing two inbred strains, followed by a few backcrosses to one of the parental strains, with subsequent inbreeding without selection.

Recombinant Inbred Strains (RI)

Inbred strains created by crossing two different inbred strains, followed by brother x sister matings for at least 20 generations. A panel of recombinant inbred strains derived from a cross between the same original parental strains can be used to establish linkage between any marker that is polymorphic between the parental strains and other polymorphic markers that have been typed in each strain in the panel.

Robertsonian Translocation (RB)

A particular type of translocation where the breakpoints in the two chromosomes occur at or near the centromere, followed by centric fusion such that the long arms now form a single metacentric chromosome with a single centromere. Any small fragments generated in the exchange are usually lost. See also Translocation.

Segregating Inbred Strain (SEG)

An inbred strain that is kept in forced heterozygosity for one or more loci. For example, the SM/J strain segregates for an agouti allele and animals are either *Aw/a* (white bellied agouti) or *a/a* (black) in color.

Sex Chromosome

Either of two chromosomes that are sexually dimorphic in species with chromosomal (as opposed to genic) sex determination. In mammals, males are the heterogametic sex, having an X chromosome and a Y chromosome, while females are the homogametic sex, having two X chromosomes.

Sperm, cryopreserved (SP)

See cryopreservation.

Spontaneous Mutation (SM)

Spontaneous mutations are those that occur in the absence of any experimental mutagenic treatment, such as irradiation or treatment with chemical mutagens.

Synonym

A synonym is an alternate symbol for a gene or allele that has appeared in the scientific literature or in public databases such as GenBank, or a formerly official symbol, withdrawn due to conform new nomenclature procedures.

Targeted Mutation (TM)

A type of mutation in which a chromosomal gene is altered by the substitution of a DNA construct assembled in vitro. These constructs are "targeted" to a specific gene and may cause deletion, insertion, or substitution in the endogenous gene. In mouse, the constructs are frequently designed to eliminate gene function; such targeted mutations are often referred to as knock outs. Some DNA constructs are designed to alter gene function through replacement of native DNA sequence; such targeted mutations are referred to as knock ins.

Transgenic Mouse (TG)

A mouse that contains stably inherited DNA which has been experimentally inserted at random into the genome. The inserted gene sequence (the transgene) may or may not be derived from mouse sequence.

Translocation (TL)

A type of chromosome rearrangement in which two nonhomologous chromosomes are each broken and then repaired in such a way that the resulting chromosomes each contain material from the other chromosome (a reciprocal translocation). See also Robertsonian Translocation.

Transposition (TP)

A type of chromosome rearrangement in which a segment of a chromosome is moved to a different location on the same chromosome or to a different chromosome, without reciprocal exchange.

Trisomy (TS)

The condition of having three chromosomes of a particular type. Trisomy for a complete extra chromosome of the standard complement is referred to as primary trisomy, such as in Down Syndrome in human where individuals are trisomic for Chromosome 21.

Unclassified Strain/Stock (UN)

Strains or Stocks that cannot be classified into any of the categories supported by the IMSR.

Wild Derived Inbred Strains (WDS) Strains formed by brother x sister matings or other inbreeding schemes from mice originally caught in the wild.