GENETIC VARIANTS AND STRAINS OF THE LABORATORY MOUSE

Second Edition

Edited by
Mary F. LYON
MRC Radiobiology Unit

and

ANTONY G. SEARLE
MRC Radiobiology Unit

for the

International Committee on Standardized Genetic Nomenclature for Mice

OXFORD UNIVERSITY PRESS · OXFORD · NEW YORK · TOKYO
GUSTAV FISCHER VERLAG · STUTTGART
1989

Oxford University Press, Walton Street, Oxford OX2 6DP
Oxford New York Toronto
Delhi Bombay Calcutta Madras Karachi
Petaling Jaya Singapore Hong Kong Tokyo
Nairobi Dar es Salaam Cape Town
Melbourne Auckland
and associated companies in
Berlin Ibadan

Oxford is a trade mark of Oxford University Press

Published in the United States
by Oxford University Press, New York

Distribution for
Continental Europe and
European Socialist Countries
by:
Gustav Fischer Verlag
Wollgrasweg 49,
D-7000 Stuttgart 70,
Federal Republic of Germany

British Library Cataloguing in Publication Data
Genetic variants and strains of the laboratory mouse—2nd ed.
1. Laboratory animals: Mice. Genetic aspects
I. Lyon, Mary F. II. Searle, Antony G.
636'.93233
ISBN 0-19-854204-6

Library of Congress Cataloging-in-Publication Data
Genetic variants and strains of the laboratory mouse/edited by M. F. Lyon and Antony G. Searle.—2nd ed.
Includes index.
1. Mice—Genetics. 2. Mice—Variation. 3. Mice as laboratory animals. 4. Mammals—Genetics. 5. Mammals—Variation. I. Lyon, M. F. (Mary F.) II. Searle, A. G. (Antony Gilbert)
QL737.R638G45 1989 599.32'32—dc 19 88–39796
ISBN 0-19-854204-6

Deutsche Bibliothek Cataloguing-in-Publication Data
Genetic variants and strains of the laboratory mouse/ed. by Mary F. Lyon and Antony G. Searle for the Internat. Committee on Standardized Genet. Nomenclature for Mice.—2. ed.—Oxford; New York; Tokyo: Oxford Univ. Press; Stuttgart: Fischer, 1989
ISBN 3-437-11268-6
NE: Lyon, Mary F. [Hrsg.]

Set by
Promenade Graphics, Cheltenham
Printed in Great Britain by
Butler & Tanner Ltd
Frome, Somerset

Extract from foreword to the first edition

This book constitutes an attempt to list and briefly describe the known genetic variants and inbred strains of the mouse, for the benefit not only of those working in mouse genetics *per se*, but also of those in diverse fields such as immunology, oncology, developmental biology, and endocrinology who now find the mouse and its variants such a valuable source of research material.

The consistent and orderly system of nomenclature used in the book for genes, chromosomal variants, and strains is the result of the work of the International Committee on Standardized Genetic Nomenclature for Mice. Mouse genetics owes a great debt to those scientists, including those to whom this book is dedicated, who realized at an early stage the need for a committee to co-ordinate genetic nomenclature for the species. The first committee, called the *Committee on Mouse Genetics Nomenclature*, was formed in 1939 when there were only 31 known gene loci and seven linkage groups, and consisted of Drs L. C. Dunn, H. Grüneberg, and G. D. Snell. They formulated the basic rules for gene nomenclature on which the present rules are founded, exhibiting and encouraging the general cooperative spirit which has been so important for the success of the nomenclature system. This was followed in 1952 by a *Committee on Standardized Nomenclature for Inbred Strains of Mice* in the formation of which G. D. Snell was again a leading figure. This committee laid down not only the rules for nomenclature of inbred strains but also the genetic criteria for the establishment of inbred strains and congenic strains. Later still, in 1958, the first two committees were merged into the present *International Committee on Standardized Genetic Nomenclature for Mice* which deals with mutant genes, inbred strains, and all types of genetic variation. G. D. Snell was its first chairman.

From the outset, the various committees realized the importance of the dissemination of information concerning nomenclature. The first step was the founding of *Mouse News Letter*, first issued in 1949 and still a major organ for the dissemination of information concerning committee rulings and nomenclature of genes and chromosomal variants. Information concerning inbred strains and their nomenclature is regularly promulgated in *Inbred Strains of Mice*, and by listings in *Cancer Research*, carried out by Joan Staats.*

Naturally, as advances in knowledge have occurred, it has been necessary to modify and extend the rules of nomenclature. This has been done by consultation between the committee members, the formation of ad hoc subcommittees, and by seeking the views of scientists concerned in the field. The new rules have been published in *Mouse News Letter* and elsewhere. Included among the items tackled in this way have been the standard karyotype of the mouse, guidelines for nomenclature of biochemical variants, rules for designation of chromosome anomalies, and nomenclature for inbred strains preserved by freezing.

M. F. Lyon
20 May, 1981

* At the time of the second edition, this information is now included in the November issue of *Mouse News Letter*.

Dedication

This book is dedicated to

LESLIE C. DUNN, HANS GRÜNEBERG, and GEORGE D. SNELL,

whose pioneering efforts produced the first rules for mouse genetic nomenclature and set an example for the habits of consultation and co-operation among mouse geneticists that still endure.

Preface to the second edition

THE years since the first edition of this book have seen not only a great increase in knowledge of mouse genetics, but also the opening up of new fields, through the application of recombinant DNA technology. In addition, there is a new interest in mouse genetics, resulting from the recent increasingly detailed knowledge of homologies between mouse and human genes and chromosomes.

Thus, the second edition includes both updated and expanded versions of the chapters in the first edition, and also six entirely new chapters. As in the first edition, the book is intended as a work of reference and provides catalogues and lists of the different classes of genetic information. The most important catalogue, that of known genes of the mouse, by Dr M. C. Green, now includes about 1500 loci. About 1000 of these are genetically mapped, and other chapters give the overall map, the data on which this is based, and some specialized maps of particular regions or types of genes. Knowledge of the correlation between the genetic map and the physical chromosomal map is becoming more detailed, both through increasing numbers of known chromosome aberrations, which provide this information, and through the technique of *in situ* hybridization. The chapter on chromosomal variants has been expanded, and there are comprehensive maps showing breakpoint positions. The normal karyotypes for G-bands, Q-bands, pachytene, and early replicating bands are also given, together with a new G-band idiogram, at a considerably higher resolution.

A new chapter shows in detail the homologies of mouse and human genetic maps, by depicting those segments of mouse chromosomes with homology to particular human chromosome segments.

As in the earlier edition, the characteristics of inbred, recombinant inbred and congenic strains are listed, in considerably expanded chapters. Restriction fragment length polymorphisms (RFLPs) are dealt with in a separate chapter giving details of the probes and the known variants. Up-to-date information on oncogenes and tumour viruses, and on the various classes of repeated DNA sequences in the mouse genome is given in other new chapters. Also new is a chapter on wild mice, which have become so important to mouse genetics, in view of the variants they carry, particularly the high frequency of RFLPs found in crosses between *Mus* species.

Finally, but importantly, the book provides three chapters on nomenclature. Use of the correct nomenclature becomes ever more essential as mouse genetics becomes more complex. Workers are urged to abide closely by the rules for nomenclature.

It is hoped that, like the first edition, this book will become a standard work of reference. Inevitably, however, new information will accrue very rapidly, and additional sources of information will be needed. Fortunately, a well organized system of information for mouse genetics exists, the main sources being as follows:

1. *Mouse News Letter*, sponsored by the International Committee on Standardized Genetic Nomenclature for Mice, is published by Oxford University Press. February and July issues are currently edited by Dr J. Peters, and the November issue by Dr M. F. W. Festing (for affiliations see list of Committee members). It publishes an annual gene list, a range of annual genetic maps, lists of chromosomal variants, lists of new gene symbols,

new linkage and gene mapping data, and lists of available DNA probes and libraries. The November issue contains details of inbred strains, congenic strains, recombinant inbred strains and their polymorphisms, and the list of substrain holders' symbols. All issues carry contributed items giving research results on mutants, linkages and other mouse genetic data.

2. *Linkage Map of the Mouse* is prepared by Dr M. T. Davisson and Dr T. H. Roderick by use of information culled from the literature and from personal communications, and is available from them.
3. The lists of DNA probes and libraries and their holders and availability, which appear in *Mouse News Letter*, are prepared by Dr J. Eppig, Jackson Laboratory, Bar Harbor, Maine 04609, USA.
4. *List of Mutations and Mutant Stocks of the Mouse* prepared by Mrs P. W. Lane, The Jackson Laboratory, contains a list of the mutant stocks, chromosome aberrations and other variants kept at the Jackson Laboratory.

The editors would like to express their gratitude to all who have helped in the preparation of this edition, particularly Mrs D. Badger, who patiently dealt with a great deal of secretarial work, and all the authors, especially Dr Margaret Green, for undertaking the heavy task of preparing the main catalogue. The editorial staff at Oxford University Press have always been ready to help us overcome the many editorial problems. Since June 1987 A. G. Searle has been helped by the award of an Emeritus Fellowship from the Leverhulme Trust.

Harwell
August 1988

M. F. LYON
A. G. SEARLE

Contributors

Mr Colin V. Beechey, MRC Radiobiology Unit, Chilton, Didcot, Oxon OX11 0RD, England

Dr Francois Bonhomme, Laboratoire de Génétique, Institut des Sciences de l'Evolution, USTL-Pl. E. Bataillon, 34060 Montpellier CEDEX, France

Dr Muriel T. Davisson, The Jackson Laboratory, 600 Main Street, Bar Harbor, Maine 04609, USA

Dr Donald P. Doolittle, The Jackson Laboratory, 600 Main Street, Bar Harbor, Maine 04609, USA

Dr Rosemary W. Elliott, Department of Molecular and Cellular Biology, Roswell Park Memorial Institute, 666 Elm Street, Buffalo, New York 14263, USA

Dr Edward P. Evans, Sir William Dunn School of Pathology, University of Oxford, South Parks Road, Oxford OX1 3RE, England

Dr Michael F. W. Festing, MRC Toxicology Unit, Woodmansterne Road, Carshalton, Surrey SM5 4EF, England

Dr Anna-Maria Frischauf, Imperial Cancer Research Fund, Lincoln's Inn Fields, London WC2A 3PX, England

Dr Margaret C. Green, 8 Seely Road, Bar Harbor, Maine 04609, USA

Dr Dorothy D. Greenhouse, Institute of Laboratory Animal Resources, National Research Council, 2101 Constitution Avenue, N.W., Washington, DC 20418, USA

Dr Jean-Louis Guénet, Unité de Génétique des Mammifères, Institut Pasteur, 28 rue du Dr Roux, 75724 Paris, CEDEX 15, France

Mr John H. Guidi, The Jackson Laboratory, 600 Main Street, Bar Harbor, Maine 04609, USA

Dr Horst Hameister, Universität Ulm, Abt. Klinische Genetik, Oberer Eselsberg, D-7900 Ulm, Germany (FRG)

Dr Nicholas D. Hastie, MRC Human Genetics Unit, Western General Hospital, Crewe Road, Edinburgh EH4 2XU, UK

Dr Norman L. Hawes, The Jackson Laboratory, 600 Main Street, Bar Harbor, Maine 04609, USA

Dr Alan L. Hillyard, The Jackson Laboratory, 600 Main Street, Bar Harbor, Maine 04609, USA

Prof Jan Klein, Department of Immunogenetics, Max Planck Institute for Biology, Corrensstrasse 42, D-7400 Tübingen 1, Germany (FRG)

Dr Christine A. Kozak, Laboratory of Molecular Microbiology, National Institute of Allergy and Infectious Diseases, 9000 Rockville Pike, Bethesda, Md 20892, USA

Mrs Priscilla W. Lane, The Jackson Laboratory, 600 Main Street, Bar Harbor, Maine 04609, USA

Dr Mary F. Lyon, MRC Radiobiology Unit, Chilton, Didcot, Oxon OX11 0RD, England

Dr Joseph H. Nadeau, The Jackson Laboratory, 600 Main Street, Bar Harbor, Maine 04609, USA

Mr Andrew H. Reiner, The Jackson Laboratory, 600 Main Street, Bar Harbor, Maine 04609, USA

Dr Thomas H. Roderick, The Jackson Laboratory, 600 Main Street, Bar Harbor, Maine 04609, USA

Dr Antony G. Searle, MRC Radiobiology Unit, Chilton, Didcot, Oxon OX11 0RD, England

Dr Imre E. Somssich, Universität Ulm, Abt. Klinische Genetik, Oberer Eselsberg, D-7900 Ulm, Germany (FRG)

Dr Benjamin A. Taylor, The Jackson Laboratory, 600 Main Street, Bar Harbor, Maine 04609, USA

Members of International Committee on Standardized Genetic Nomenclature for Mice

M. F. LYON
MRC Radiobiology Unit, Chilton, Didcot, Oxon OX11 0RD, England (Chairman)

D. W. BAILEY
The Jackson Laboratory, Bar Harbour, Maine 04609, USA

M. T. DAVISSON
The Jackson Laboratory, Bar Harbor, Maine 04609, USA

P. DÉMANT
Netherlands Cancer Institute, Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands (Editor, *Mouse Membrane Alloantigen News*)

R. W. ELLIOTT
Roswell Park Memorial Institute, 666 Elm Street, Buffalo, New York 14263, USA

M. F. W. FESTING
MRC Toxicology Unit, Woodmansterne Road, Carshalton, Surrey SM5 4EF, UK (ICLAS Representative) (Editor, *Inbred Strains Issue*)

A.-M. FRISCHAUF
Imperial Cancer Research Fund, Lincoln's Inn Fields, London WC2A 3PX, UK

J. FOREJT
Institute of Molecular Genetics, Czechoslovak Academy of Sciences, Videnska 1083, 142 20 Praha 4, Czechoslovakia

J.-L. GUÉNET
Unité de Génétique des Mammifères, Institut Pasteur, 28 rue du Dr Roux, 75724 Paris, CEDEX 15, France

R. S. HOLMES
Griffith University, Nathan, Brisbane, Queensland 4111, Australia

J. KLEIN
Department of Immunogenetics, Max Planck Institute of Biology, Corrensstrasse 42, D-7400 Tübingen 1, Germany (FRG)

K. MORIWAKI
Department of Cytogenetics, National Institute of Genetics, Misima, Sizuoka-ken, 411 Japan

H. C. MORSE
National Institute of Allergy and Infectious Disease, Bldg 7, Rm 383, NIH, Bethesda, Maryland 20205, USA

J. PETERS
MRC Radiobiology Unit, Chilton, Didcot, Oxon OX11 0RD, England (Editor, *Mouse News Letter*)

T. H. RODERICK
The Jackson Laboratory, Bar Harbor, Maine 04609, USA

L. B. RUSSELL
Biology Division, Oak Ridge National Laboratory, P.O. Box Y, Oak Ridge, Tennessee 37831, USA

A. G. SEARLE
MRC Radiobiology Unit, Chilton, Didcot, Oxon OX11 0RD, England

J. STAATS
The Jackson Laboratory, Bar Harbor, Maine 04609, USA

Contents

1 RULES AND GUIDELINES FOR GENE NOMENCLATURE

COMMITTEE ON STANDARDIZED GENETIC NOMENCLATURE FOR MICE
Chairman: MARY F. LYON

These rules were adopted by the Committee in 1984. They are the most recent version of rules first published by Dunn, Grüneberg, and Snell in 1940 (3), revised by the Committee in 1963 (1), amplified for biochemical variants by the Committee in 1973 (2), and revised again in 1979 (5).

1.1 Rules for gene nomenclature

1.1.1 Names of gene loci

Names of gene loci should be brief and should be chosen so as to convey as accurately as possible the character by which the gene is usually recognized. Such a character may range from a coat colour or morphological effect to change in an enzyme or other protein, disease susceptibility or resistance, resemblance to a human syndrome, or a DNA sequence identified by a DNA probe for the gene.

1.1.2 Symbols for gene loci

Symbols for gene loci should typically be two-, three-, or four-letter abbreviations of the name. For convenience in alphabetical listings, the initial letters of names and symbols should where possible be the same. Arabic numbers may be included for proteins in which a number is part of the recognized name or abbreviation, but the symbol should always begin with a letter, e.g. *G6pd*, glucose-6-phosphate dehydrogenase; *B2m*, β2-microglobulin; *C3* and *C4*, third and fourth components of complement, respectively. Roman numbers and Greek letters should not be used. Names of persons or places are in general not suitable for gene names or symbols.

Except in the case of loci first discovered because of a recessive mutation (see Section 1.1.5), the initial letter of the locus symbol should be a capital, and all others lower case.

In published articles gene symbols should be set in italics, e.g. *dw*, dwarf; *Hbb*, haemoglobin β-chain.

Identification of new loci should not be assumed from the discovery of variation, whether morphological, biochemical, or antigenic. Appropriate genetic tests should be made to show Mendelian segregation, and identity or not with known loci should be established as far as possible by mapping or tests for allelism. Loci may also be identified by somatic cell genetics, or studies of DNA.

A proposed new symbol must not duplicate one already used for another locus, even if the gene effect is very different. Listing of a gene symbol in *Mouse News Letter* establishes priority.

1.1.3 Loci which are members of a series

Loci which are members of a series specifying similar proteins or other characters (e.g. isoenzymes, lymphocyte antigens, histocompatibility antigens) should be designated by the same letter symbol with the addition of a hyphen and a distinguishing number: e.g. *H-1*, *H-2*, etc., histocompatibility loci; *Es-1*, *Es-2*, etc., for esterase loci.

For morphological or 'visible' loci with similar effects (e.g. 'waltzing' genes, hair-waving genes) distinctive names should be given since the gene actions and gene products may ultimately prove to be very different, e.g. *v*, waltzer, and *kr* kreisler; *Sl*, steel, and *W*, dominant spotting.

1.1.4 Homology with other organisms

It is highly desirable that terminology for homologous genes should be standardized among species. Therefore, when choosing a gene symbol an attempt should first be made to discover and use any symbols already adopted for this gene in other species. However, care should be taken that such symbols *do not duplicate any already in use in the mouse* for other loci. If duplication would occur, then the symbol should be modified to one resembling that used in the other species but not duplicating that used for a different gene in that (or

another) species, e.g. carbonic anhydrase: man *CA*, mouse, *Car-1* and *Car-2*; catalase: man *CAT*, mouse, *Cas-1*.

Where possible the numbering of homologous loci in a series should be made concordant in various species, with locus *1* in the mouse corresponding to the locus *A* in the other species, locus *2* with locus *B*, and so on.

1.1.5 Alleles

Alleles should be designated by the locus symbol with an added superscript (in italics when printed). In computerized symbols the superscript may be denoted by prefixing an asterisk, e.g. Hbb^d or *Hbb*d*.

1. Exceptions occur for the first discovered allele in those cases in which there is clearly a wild type. No superscript is then used, e.g. *nu*, nude; *Ca*, caracul. When further alleles are discovered the first mutant allele may still be without a superscript.

2. Recessive alleles should be indicated by the use of a lower case initial letter for a mutant gene, e.g. *a*, nonagouti; *nu*, nude.

All other alleles, whether dominant, codominant, or having dominance relationships which vary with method of assessment, should be indicated by the use of a capital initial letter followed by lower case letters, as in the locus symbol, e.g. *Re*, rex; *Ta*, tabby.

3. Allele superscripts should typically be one or two lower case letters and, if possible, should convey additional information about the allele, e.g. c^{ch}, chinchilla allele of *c* or albino; Mi^{wh}, white allele at the microphthalmia locus; Hbb^d, haemoglobin β-chain allele giving a *diffuse* band after electrophoresis. If information is too complex to be conveyed conveniently in the symbol (e.g. biochemical properties, antigenic specificities), the alleles are given superscripts and the information concerning the allelic properties is shown in catalogues or tables, e.g. $Pgm\text{-}1^a$, $Pgm\text{-}1^b$; $H\text{-}2^a$, $H\text{-}2^b$, etc.

4. Wild type alleles should be designated by a + sign, with the locus symbol as a superscript e.g. $+^d$, $+^c$. Reversions from a mutant allele to wild type should be distinguished from the original wild type allele by designating them by the locus symbol, with a + sign as superscript e.g. d^+, pe^+. A + sign only may be used when the context leaves no doubt as to the locus represented e.g. in genetic formulae.

5. Indistinguishable alleles of independent origin (e.g. reoccurrences, reversions to wild type) should be designated by the existing gene symbol with a series symbol (see below) appended as a superscript in italics. If the gene symbol already has a superscript, this should be separated from the series symbol by a hyphen. The series symbol should consist of an Arabic numeral corresponding to the serial number of the variant in any given laboratory, plus an abbreviation indicating the discoverer or laboratory of origin. Where an abbreviation for the designation of inbred substrains or sublines has already been assigned, this should be used. Otherwise, the abbreviations used should follow the rules for substrain designation and not duplicate an existing symbol in the standard list of abbreviations. To avoid the confusion of the numeral *1* and the letter *l*, a first-discovered variant may be left unnumbered, and the second variant numbered *2*. Examples: c^{4Rl}, the fourth reoccurrence of *c* found by Russell; $a^{t\text{-}7J}$, the seventh reoccurrence of a^t found at the Jackson Laboratory; d^{+J} and d^{+2J}, the first and second reversions from *d* to d^+ found at the Jackson Laboratory.

6. Mutations or other variations occurring in known alleles may be denoted by a superscript *m* followed by an appropriate series symbol (as above) and separate from the original allele symbol by a hyphen; e.g. $Mod\text{-}1^{a\text{-}m1Lws}$, the first mutant allele of $Mod\text{-}1^a$ found by Lewis.

For known deletions of all or part of an allele the superscript *m* may be replaced with *dl*.

Information on the allele of origin of mutations may be valuable in elucidating changes in DNA sequence.

1.1.6 Phenotype symbols

Phenotype symbols, where these are necessary (e.g. antigen loci, enzyme loci), should be the same as genotype symbols except that symbols for phenotypes should be in capitals, not italicized, and with superscripts lowered to the line. The phenotypes of heterozygotes should be written as in the following example: GPI-1A, GPI-1B, and GPI-1AB are phenotypes associated with the *Gpi-1* locus.

In those cases in which information concerning the subunit structure of a protein is available, phenotype symbols should reflect the subunit composition. Where multiple loci exist for a single enzyme (genetic isozymes), locus *1* should be considered to synthesize the A subunit, locus *2* the B subunit, etc. Phenotypes are then designated according to the rules of the International Union of Biochemistry (6), e.g.

Locus	Isozyme phenotype
Pgm-1	PGM-A
Pgm-2	PGM-B
Adh-1	ADH-A_2

Adh-2	$ADH\text{-}B_2$
Ldh-1	$LDH\text{-}A_4$

Phenotype symbols for allelic variants reflect the subunit structure; i.e. allele 1^a synthesizes the A^a subunit; allele 1^b the A^b subunit, etc.; e.g.

Genotype	Phenotype
$Ldh\text{-}1^a/Ldh\text{-}1^a$	$LDH\text{-}A_4^a$
$Ldh\text{-}1^b/Ldh\text{-}1^b$	$LDH\text{-}A_4^a$
$Ldh\text{-}1^a/Ldh\text{-}1^b$	$LDH\text{-}A_4^aA$, $A_3^aA^b$, $A_2^aA_2^b$, $A^aA_3^b$, A_4^b

1.1.7 Gene complexes

Gene complexes are considered to exist when a number of apparently functionally related loci are genetically closely linked. Alternative states of complexes are referred to as *haplotypes* rather than alleles.

Known complexes are of two main types: (a) less extensive complexes involving duplicate loci or in which operators or *cis*-acting regulators of structural genes for protein show little or no recombination with the loci on which they act: and (b) very extensive complexes, possibly involving hundreds of related loci, for which special rules may be necessary. The *H-2* and the immunoglobin complexes are in category (b).

Less extensive complexes involving operators, cis-acting regulators, or duplicate loci

The existence of a gene complex, rather than multiple types of variation in a structural gene, should not be postulated unless there is good evidence. It should be remembered that different mutations in a structural gene may affect not only electrophoretic mobility but also activity and stability. In addition, changes in 5′ or 3′ regulatory sequences may cause apparent changes in tissue specificity or inducibility. Thus, such changes in effect should be attributed to mutations in the structural gene unless there is good evidence otherwise.

To distinguish different loci of a complex, the basic symbol should have appended a single lower case letter designating the presumed function or means of identification of the locus, such as *s* (structural), *e* (electrophoretic), *r* (regulatory), or *t* (temporal). This letter should be set off by a hyphen, as in *Bgl-e* (beta-galactosidase electrophoretic), except in the case of numbered unlinked loci in a series, in which case the letter should follow the number without a second hyphen, as in *Adh-3t*, (alcohol dehydrogenase-3 temporal). When it is discovered that a previously described locus is part of a complex, a letter indicative of its function or means of identification should be added to the basic symbol to represent the already known locus, as well as a different letter for the newly discovered locus. An example is: the *Adh-3* locus, after discovery of a temporal regulator, becomes *Adh-3e* (*Adh-3* electrophoretic) and the regulator is called *Adh-3t* (*Adh-3* temporal). The basic symbol then represents the entire complex; if necessary for clarity, the complex may be additionally indicated by enclosing the basic symbol in parentheses or in brackets. Haplotypes are designated by the symbol for the complex with a superscript small letter. The components of the haplotype can be briefly indicated as in the following example: $Adh\text{-}3^a$ or $(Adh\text{-}3)^a = Adh\text{-}3e^a$ $Adh\text{-}3t^a = Adh\text{-}3e^at^a$;. In the case where two or more closely linked and functionally related structural loci have been given serial numbers, the complex, loci, and haplotypes should be indicated as in the following example: complex, *Amy* or (*Amy*); loci, *Amy-1*, *Amy-2*; haplotypes, Amy^a or $(Amy)^a = Amy\text{-}1^a$ $Amy\text{-}2^a = Amy\text{-}1^a2^a$.

Distantly acting regulators should be given locus symbols different from but related to the locus they regulate and preferably with the same initial letter, for example, *Ldr* (a regulator of the lactate dehydrogenase locus *Ldh*).

Suffixes so far used, either for loci in gene complexes, or for subdividing series of loci, include:

-C Constant region (immunoglobulins),
-e Electrophoretic,
-m Mitochondrial,
-r Regulatory,
-s Structural,
-t Temporal,
-V Variable region (immunoglobulins).

Very extensive complexes with special rules

The list of extensive complexes with special rules continually increases. The need for special rules for each arises because the various complexes differ widely in their structure and no suitable single nomenclature system has yet been found that is adequate for all these complexes.

Some complexes with special rules include:

(a) *H-2* complex;

(b) immunoglobulin complexes;

(c) globin gene complexes;

(d) homeobox-containing gene complexes;

(e) *t*-complex.

(a) *The histocompatibility-2 or H-2 region complex.* Symbols for haplotypes should be superscripts, consisting of small letters and Arabic numerals, assigned according to the standing rules for the *H-2* complex (7), such as $H\text{-}2^b$, $H\text{-}2^{bm1}$, $H\text{-}2^{ap4}$. Parts of the complex

that are shown to be separable from each other by recombination are referred to as *regions*, and may consist of many loci. The regions are denoted by symbols on the line in capital letters, for example, *K*, *I*, *S*, and *D*. In referring to alternative states or 'alleles' of particular regions, the symbol of the region is appended by a superscript denoting the haplotype from which the region originates, for example K^k, S^d, D^b. The individual loci have symbols reflecting their characteristics, which are chosen according to the valid 'Rules for gene nomenclature in mice', the rules for nomenclature of the *H-2* gene complex (7), and rules for the nomenclature of the *I* region (13). Examples: *H-2K*, *Ia-1*, *C4*, *H-2L*. The allelic forms of these loci are designated by superscript letters denoting the haplotype of origin, for example, $H\text{-}2K^b$, $C4^d$. Rules for nomenclature of mutant haplotypes were summarized by Kohn *et al.* (8) and by Shreffler (14).

(b) *The immunoglobin complexes.* The heavy-, kappa-, and lambda-chain regions are designated *Igh*, *Igk*, and *Igl*, respectively. Haplotypes are designated by the symbol for the region with a superscript lower case letter. The constant subregions are designated *Igh-C*, *Igk-C*, and *Igl-C*, and individual loci in these subregions are designated by Arabic numbers following the hyphen and assigned chronologically, for example, *Igh-1*, *Igh-2*, etc., *Igl-1*. Allelic symbols are superscript lower case letters denoting the haplotype, or one of the haplotypes, in which the allele occurs. The variable subregions are designated *Igh-V*, *Igk-V*, and *Igl-V*, and individual loci in these subregions are designated by a two- or three-letter symbol or by two letters and a number following the hyphen, for example, *Igh-Dex*, *Igh-Pc*, *Igk-Ef1*. The symbol following the hyphen for the variable-region loci should be related to the antigen for which the immunoglobulin is specific or to the method used for recognizing the variant. Allelic symbols are superscript lower case letters such as: *a* and *b* in the case where allelic markers are well established, for example, $Igk\text{-}Ef1^a$ and $Igk\text{-}Ef1^b$ or *a* and *o* in the case where the allelic nature of markers is in doubt and the alleles are postulated to determine a marker and its absence, for example, $Igh\text{-}Dex^a$ and $Igh\text{-}Dex^o$. These rules are given in greater detail by Green (4).

The nomenclature systems for (c) globin gene complexes and (d) homeobox-containing gene complexes are summarized at the end of these rules.

1.1.8 Pseudogenes

Pseudogenes located away from the main gene complex should be denoted by the suffix *ps*, separated from the locus symbol by a hyphen, and followed by an appropriate serial number; e.g. *Hba-ps3*, *Hba-ps4*, pseudogenes of α-globin located away from the *Hba* complex.

1.1.9 Lethals

Appropriate locus symbols for recessive lethals with no known heterozygous effect and unidentified function consist of a lower case letter *l* followed by the chromosome number of location in parentheses, and by a hyphen and series symbol indicating the serial number of the lethal in the laboratory of origin, e.g. *l(5)-1Rk*, the first lethal on chr. 5 found by Roderick; *l(17)-2Pas*, the second lethal on chr. 17 found at the Pasteur Institute.

Such symbols should be considered as provisional. The lethal should be renamed if found to be allelic with a known gene, or if the underlying defect becomes understood.

1.1.10 Viruses

Nomenclature for genes related to the expression of viral antigens, or to sensitivity or resistance to viruses, should follow the standard rules for gene nomenclature, i.e. symbols should be italicized, with the initial letter a capital and all others lower case. Where possible and appropriate, the letters of the symbol should be those by which the virus is usually known; e.g. *Mtv-1*, a locus concerned in induction of mammary tumour virus, MTV. Successive loci concerned with the same virus should be distinguished by appending a number separated by a hyphen; e.g. *Fv-1*, *Fv-2*, loci concerned with resistance to Friend virus.

Locus symbols ending in *v* should be reserved for virological loci.

1.1.11 Oncogenes

Nomenclature for mouse cellular oncogene sequences should follow the standard nomenclature for oncogenes. However in lists of symbols and maps the prefix *c-* denoting cellular sequence should be omitted and the initial letter of the symbol should be capitalized; e.g. *c-myc* becomes *Myc*, myelocytoma oncogene; *c-Hras-1* becomes *Hras-1*, Harvey rat sarcoma-1 oncogene; *c-erba* becomes *Erba*, avian erythroblastosis oncogene.

The names and symbols of oncogenes should be regarded as provisional until the true functions of the genes become known, when they should be renamed, e.g. *Erbb* becomes epidermal growth factor receptor, *Egfr*; *Sis* becomes platelet derived growth factor, beta polypeptide, *Pdgfb*.

1.1.12 Mitochondrial genome

Loci in the mitochondrial genome should be denoted by the prefix *mt-* set off from the main symbol by a hyphen.

1.1.13 Restriction fragment length polymorphisms

There are a number of different situations in which restriction fragment length polymorphism may occur.

(a) variation in DNA sequence within exons of a known gene;

(b) variation in DNA sequence within introns and/or flanking sequences of a known gene;

(c) variation in DNA sequence outside situations in (a) or (b) but detected by a probe for the known gene, e.g. the *Hpa* site variation 5 kb from the 3′ end of the human β-globin structural gene;

(d) variation in DNA sequence detected using an arbitrary DNA sequence as a probe.

The type of variation detected in (a) and (b) should be described using current rules for nomenclature of gene loci and alleles. Thus these variants could be listed both in a compilation of restriction fragment length variants and in lists of gene loci.

For (c) symbols for the restriction fragments should begin with *D* (for DNA), followed by the gene symbol, followed by a number, e.g. the *Hpa* site variant cited above would be symbolized *DHbb1*, the 1 indicating that this was the first probe found. The variation in possession of the *Hpa* site can be described in terms of 'alleles'. Thus the presence of the site would be designated $DHbb1^a$ and the absence $DHbb1^b$. If the allele in which the variation occurs is known this should be indicated in the symbol; e.g. $DHbb^{d1a}$.

For (d) it is not possible to ascertain if the variation fits into categories (a), (b) or (c). The nomenclature should follow that in human gene mapping for provisional nomenclature (15).

An arbitrary probe is given a name composed of four parts:

(i) D for DNA.

(ii) 0, 1 . . . . . 19, XY for the chromosomal assignment, with 0 for unassigned segments.

(iii) a symbol indicating the laboratory or scientist describing the probe. This should be the abbreviation used to designate inbred substrains, e.g. Pas for Pasteur Institute.

(iv) a number to give uniqueness to the probe; e.g. *D1Pas5*, the fifth chromosome 1 probe, developed at the Pasteur Institute; *D17Leh48*, a chromosome 17 probe designated no. 48 by Lehrach.

Alleles of arbitrary DNA sequences could be designated as follows: $D1Pas5^a$ could indicate possession of a restriction site for a particular enzyme and $D1Pas5^b$ its absence.

If the arbitrary sequence is later identified to be at a known locus, the nomenclature should be altered to take this into account.

Anonymous DNA segments from the human genome which hybridize with mouse DNA and are mapped to a mouse chromosome should retain their human symbol, and this should be followed by a lower case *h* to denote the human origin e.g. *D21S56h*, a DNA segment from human chromosome 21 (which hybridizes to mouse chromosome 17).

1.1.14 Antigenic variants

Symbols adopted for loci concerned in cell-membrane alloantigens should be based on the method of demonstrating such loci:

1. Loci primarily demonstrable by transplantation techniques should be designated by an initial *H*, e.g. *H-1*, *H-2*, etc.

2. Loci demonstrable by red-cell agglutination should be designated by the letters *Ea*, e.g. *Ea-1*, *Ea-2*, etc.

3. Loci of genes coding for a cell surface molecule on lymphocytes, or shared by lymphocytes and other cell types, and detected by serological or biochemical methods, and for which there is a demonstrable polymorphism, should be designated by the letter *Ly*, e.g. *Ly-4*, *Ly-5*, etc. More details of nomenclature for *Ly* antigen genes are given at the end of these rules.

4. Similarly, other loci involving other cell types should be denoted by symbols indicating the cell type, e.g. *Pca*, plasma cell antigen; *Tla*, thymus leukaemia antigen.

Appropriate genetic nomenclature for the H-2 complex has been put forward by Klein *et al.* (7), and for the *I* region of this complex by Shreffler *et al.* (13); this nomenclature should be followed.

A registry of symbols for *H-2* haplotypes and antigenic specificities is maintained and additions to the

register are listed in Mouse Membrane Alloantigen News (MMAN) incorporated in *Mouse News Letter*. New variants should be registered with the editor of MMAN, at present Dr P. Demant.

1.2 Nomenclature for special classes of genes and gene complexes

1.2.1 Guidelines for nomenclature of biochemical variants

Biochemical nomenclature should be in accord with the rules of the International Union of Biochemistry, Commission on Biochemical Nomenclature. The nomenclature recommended by the Commission is published periodically in major international biochemical journals, such as the *Journal of Biological Chemistry* and the *Biochemical Journal*. Enzymes and other biochemicals have both trivial and formal names. The correct formal name should be given the first time a substance is mentioned in a publication; trivial or abbreviated names can be used subsequently. Example: G6PD, GPD, or Gd, abbreviations of glucose-6-phosphate dehydrogenase (E.C. 1.1.1.49); D-glucose 6-phosphate:NADP oxidoreductase). This nomenclature is used in periodicals, reference works, and textbooks of biochemistry.

Symbols of structural loci should typically be two-, three-, or four-letter abbreviations (italic type) of the official Commission name of the enzyme, protein, or other substance affected. The initial letter of the symbol should be capitalized. Example: *Gpi-1*, the first identified structural locus of glucosephosphate isomerase (E.C. 5.3.1.9; D-glucose 6-phosphate ketol-isomerase). In the case of biochemical variants, the use of the locus symbol with a lower case initial letter to indicate recessive mutant genes and with a capital initial letter to indicate dominant mutant genes should generally be avoided. Such nomenclature is not suited to polymorphic systems of alleles, and the dominance–recessive relationship usually varies and depends on the method used to assess it.

Greek letters preceding the name of an enzyme or other protein, should be changed to an appropriate English letter and placed at the end of the locus symbol; e.g. *Fuca*, α-fucosidase. This permits a rational alphabetic ordering of locus symbols.

Similarly, adjectives describing the tissue specificity or other property of an enzyme or protein should normally be placed after the noun, again in order to allow appropriate alphabetic ordering of symbols, e.g. *Actc*, actin, cardiac; *Acts*, actin, skeletal.

A series of loci specifying the structure of isoenzymes that catalyse the same or similar reactions but are structurally different, or the different polypeptide chains of a protein, can be designated by the same letter symbol for the structural locus with the addition of a hyphen and a distinguishing number. Example: *Pgm-1* and *Pgm-2*, loci of structurally different isoenzymes of phosphoglucomutase (E.C. 2.7.5.1; phosphotransferase).

Homology with other organisms. It is highly desirable that terminology for homologous genes should be standardized among species. Therefore, as in the standard rules, when choosing a gene symbol an attempt should first be made to discover and use any symbols already adopted for this locus in other species. However, care should be taken that such symbols *do not duplicate any already in use in the mouse* for other loci. If duplication would occur, then the symbol should be modified to one resembling that used in the other species but not duplicating that used for a different gene in that (or another) species, e.g. carbonic anhydrase: man CA; mouse, *Car-1* and *Car-2*.

Where possible the numbering of homologous loci in a series should be made concordant in various species, with locus *1* in the mouse corresponding to the locus *A* in other species, locus *2* with locus *B*, and so on.

It is not appropriate to insert the letter m or M (for mouse) as the first letter of the symbol for a locus with homologues in other species since this would lead to all mouse locus symbols beginning with the same letter.

Alleles should be designated by the locus symbol with an added superscript as in the standard rules. In describing alleles, whether found in inbred strains or in the wild, it is desirable that the phenotype of a number of widely used inbred strains be reported. One strain should arbitrarily be designated the prototype strain for each allele, since variation that has not been detected by the methods used may be present within each allelic class. If an apparently identical allele in other strains is found by new methods to be different from that in the prototype strain, it should be assigned a new alphabetical symbol as a superscript and a prototype strain designated. This system permits the orderly assignment of symbols to newly identified alleles and allows ready comparisons of new variants with previously reported variants.

Locus and allele symbols are necessarily brief and cannot contain more than a small fraction of the known information. Additional information may be contained in gene descriptions which in some cases can be collected in catalogues or tables. The haemoglobin α-