Molec. gen. Genet. 175, 129–133 (1979)

MGG

A Uniform Genetic Nomenclature for the Nematode *Caenorhabditis elegans*

H. Robert Horvitz[1], Sydney Brenner[2], Jonathan Hodgkin[2], and Robert K. Herman[3]

[1] Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[2] Medical Research Council Laboratory of Molecular Biology, Cambridge CB2 2QH, England
[3] Department of Genetics and Cell Biology, University of Minnesota, St. Paul, MN 55108, USA

Summary. A uniform system of genetic nomenclature for the nematode *Caenorhabditis elegans* is described. Convenient ways are specified to designate genes, mutations and strains, and to attempt to avoid name duplications.

An extensive genetic characterization of the small soil nematode *Caenorhabditis elegans* was reported by Brenner in 1974. Continued genetic interest in this organism seems likely; at present, twenty-three laboratories are engaged in genetic studies of *C. elegans* (Appendix A). Representatives from all of these laboratories have agreed upon a uniform system of genetic nomenclature, which is described below. These general recommendations should be considered as flexible guidelines, to be followed when possible and extended and modified when necessary.

The system is based upon the nomenclature introduced by Brenner (1974). It is simple, clear, convenient, and uniform. The system is easy to use both in the laboratory and in typed and printed texts (e.g., superscripts, subscripts, and Greek letters are avoided). By establishing centers at which gene, mutation, and strain names are registered, it prevents duplicate assignments of particular names. Furthermore, the system should be readily understood by workers in other biological fields. Similar goals have led to the standardization of genetic nomenclature for mice (Committee on standardized nomenclature for mice, 1963), *Neurospora* (Barratt and Perkins, 1965), bacteria (Demerec et al., 1966), yeast (von Borstel, 1969), and *Aspergillus* (Clutterbuck, 1973).

Almost all *C. elegans* mutants have been derived from the Bristol strain designated N2 (Brenner, 1974). This strain is called wild-type. A major purpose of the system of nomenclature described below is to allow one to define the known genetic differences between a given *C. elegans* strain and N2.

For offprints contact: H.R. Horvitz

I. Gene Names

The existence of a gene is revealed by the phenotypic effect(s) of a mutation. In general, mutations that have similar phenotypic effects, fail to complement, and are located in virtually identical positions on the genetic map are assumed to be allelic (i.e., to lie in the same gene). However, it is well known that alleles do not always fulfil all of these criteria; thus, each case must be evaluated individually. More detailed analyses can rely on a more precise definition of a gene as the DNA base sequence coding for a given polypeptide chain, or, in some cases (e.g. tRNA's or rRNA's), for a given RNA molecule; *cis*-acting control elements may also be included.

In *C. elegans* genetics, gene names usually refer to relatively broad phenotypic categories. Genes within such a category share a general name, which consists of three italicized (or underlined, in typed copy) lower case letters, e.g. *unc* or *dpy*. This general name refers to the phenotype originally detected and/or most easily scored. Pronounceable three letter names are most convenient. Ideally, the gene name should be general enough to avoid ambiguities when additional alleles are identified. For example, genes determining drug sensitivity are named after the drug and not after the phenotype conferred by particular alleles (i.e., resistance, hypersensitivity, or dependence); thus, the *lan* and *lev* genes determine lannate and levamisole sensitivity, respectively. Similarly, whether particular alleles are dominant or recessive is not indicated in a gene name; upper case letters are not used in gene names to refer to dominant alleles. The general gene names currently in use are listed in Appendix B.

0026-8925/79/0175/0129/$01.00

Different genes within the same general category are distinguished by different italicized Arabic numbers, which are separated from the general name by a hyphen (-). For example, *unc-15* and *unc-54* are different genes. When appropriate, gene names can include an italicized Roman numeral (*I, II, III, IV, V, X*) to indicate the linkage group on which that gene maps, e.g. *dpy-18 III* or *lon-2 X*.

To avoid using the same name for different genes, each three letter code is assigned by only one laboratory (Appendix B). Other laboratories can either a) use an established general gene name and send for appropriate numeric gene assignments, or b) designate a new general name for that phenotype and make assignments. When practical, the former alternative is preferable. New gene names should be assigned only if genetic analysis has established nonallelism with the known genes of similar map position and phenotypic effect.

Occasionally, it may be discovered that two gene names have been assigned to a single gene. If the general gene names are the same, the name with the lower numerical designation should be maintained; hence, two mutations originally assigned to *unc-60* and *unc-66* now both are considered to be alleles of *unc-60*. In other instances, the gene name with the higher priority is that a) which has had the greater number of designated alleles, b) with the more common general name, c) suggesting the more easily scored phenotype, or d) published first. These criteria generally should be applied in the order listed.

All three letter general gene names should be registered with the central clearing house, which is presently administered by Dr. R. Herman, Department of Genetics and Cell Biology, University of Minnesota, St. Paul, MN 55108.

II. Mutation Names

A mutation confers a heritable change in phenotype. Every mutation should have a unique designation. Generally, mutation names consist of 1 or 2 lower case italicized letters followed by an italicized Arabic number e.g. *e128* or *mn103*. Each letter prefix is used by only one laboratory (Appendix A) and should be registered with the central clearing house. To avoid confusion with the numbers, letter prefixes should not end in "i" or "l" or "o". Mutations initially recognized to be chromosomal aberrations are named somewhat differently (see below).

Suffixes indicating particular characteristics of a mutation can follow the mutation name. These descriptive suffixes should consist of lower case nonitalicized letters and should be explicitly defined when first used. To avoid confusion with gene names, three letter descriptive suffixes should be avoided. Descriptive suffixes should not start with "l" or "o". Useful suffixes include: dm, dominant; r, resistant; rl, recessive lethal; s, sensitive; sd, semidominant; ts, temperature sensitive. Thus, *e1348*rl and *e184*sd are mutation names. Multiple descriptive suffixes are separated by commas, e.g. *e1157*ts, sd. The presence of a descriptive suffix is for convenience and does not further define the mutation involved; i.e. *e1324*ts and *e1324* refer to the same mutation.

When gene and mutation names are used together, the mutation name is included in parentheses after the gene name, e.g. *unc-4(e120) II* or *dpy-11(e224)*. When unambiguous, gene names should be cited in preference to mutation names. For example, in a manuscript in which *e190* is the only mutation in *unc-54 I* mentioned, the name *unc-54(e190) I* should be listed early in the paper and thereafter referred to as *unc-54*. If convenient, a suffix describing the mutation used can be included in parentheses. For example, once defined as *dpy-13(e184*sd*) IV*, this mutation can be referred ot as *dpy-13(*sd*)*.

Wild-type alleles either are not indicated or are designated by a plus (+) sign. When unambiguous, a plus sign can be used alone, e.g. *rol-2/+*. Alternatively, the wild-type allele can be designated with a plus sign after the gene name: *sma-2+* or *sma-2+ III*. The plus sign follows directly on the same line. It is also acceptable to designate the wild-type allele according to the convention for mutations described above, i.e. by including the plus sign in parentheses: *sma-2(+)* or *sma-2(+) III*.

III. Multiple Mutations

Mutants carrying more than one mutation on a given linkage group are designated by sequentially listing mutations or mutant genes according to their left-to-right order (if known) on the standard genetic map (Brenner, 1974). Mutations or mutant genes on different linkage groups are separated by a semi-colon (;) and listed in the linkage group order I, II, III, IV, V, X, e.g. *dpy-5 him-1 I; unc-17 IV; nuc-1 X* or *e61 e879; e245;e1392* or *dpy-5(e61) him-1(e879) I; unc-17(e245) IV; nuc-1(e1392) X*.

Heterozygotes, with allelic differences between two homologous chromosomes, are designated by separating mutations or mutant genes on the two chromosomes with a slash (/), e.g. *e189/e364* or *unc-32/dpy-18*. If only one of two homologous chromosomes is mutant, one or more plus signs designate the wild-type chromosome, e.g. *dpy-13 IV/+* or *dpy-10 unc-52/+ +*. If both chromosomes are mutant,

plus signs are optional; thus, *e128/e120* is equivalent to *e128 +/+ e120*. If two homologous chromosomes are identical, mutations or genes need be listed only once; if they differ, complete genotypes are designated for both chromosomes (plus signs again are optional). Thus, *dpy-5 him-1 +/+ him-1 unc-15; lon-1* is homozygous for *him-1 I* and *lon-1 III* and heterozygous for *dpy-5 I* and *unc-15 I*.

The absence of a second X chromosome in hemizygous males generally is implicit, but if convenient can be specified using an "0", e.g. *lon-2/0*. Trisomics are designated using two slashes to separate the three homologous chromosomes, e.g. *lon-2 X/dpy-6 X/unc-19 X*. Other aneuploids would be indicated similarly. Polyploids are designated by listing in italics the number of haploid sets of autosomes (abbreviated "A"), a semi-colon, and the number of sex chromosomes ("X"); e.g., *4A;4X* is a tetraploid with four sets of autosomes and four sex chromosomes.

IV. Chromosomal Aberrations

Six elementary types of chromosomal aberrations have been observed in genetic studies of *Drosophila melanogaster* (Lindsley and Grell, 1968), and it is anticipated that similar rearrangements will be identified in *C. elegans* (e.g., Herman et al., 1976; Herman, 1978). These basic types of rearrangements are translocations, ring-chromosomes, inversions, transpositions, duplications, and deficiencies; they are abbreviated T, R, In, Tp, Dp, and Df, respectively. Definitions of these chromosomal aberrations are presented by Lindsley and Grell (1968). In *C. elegans* genetics, deficiencies at present must be defined genetically rather than cytologically.

Chromosomal aberrations are named by listing first the laboratory mutation name prefix (Appendix A); one of the six abbreviations listed above; an Arabic numerical designation; optionally, the affected linkage group(s) in parentheses (a lower case "f" indicates a free or fragment chromosome); and finally, also optionally, affected gene(s). As in all genotype designations, chromosomal aberrations are italicized. Examples of names that correspond to this format are *mnDp2* and *sDf3(I)dpy-5*. For translocations, linkage groups are listed in the order I, II, III, IV, V, X as in *mnT1(II;X)*. For duplications, the original linkage group of the duplicated segment precedes the recipient linkage group; thus, *mnDp1 (X;V)* indicates that a segment of the X chromosome has become linked to linkage group V. Similarly, *eDp6 (III;f)* names an aberration in which a segment of linkage group III behaves as a fragment chromosome.

These various categories of chromosomal aberrations are not mutually exclusive. The name used should be the first appropriate in the order T, R, In, Tp, Dp, Df. Once a name for a chromosomal aberration has been published, this name should not be altered, even if further characterization indicates that a higher category is more appropriate.

V. Strain Names

A strain is a set of individuals of a particular genotype with a capacity to produce additional individuals of identical genotype. Strains can and should be stored in liquid nitrogen (Brenner, 1974). The designation of each strain should be unique. Strain names consist of 2 or 3 non-italicized upper case letters followed by an Arabic number. Strain name letter prefixes refer to the laboratory of origin (Appendix A) and should be different from mutation name letter prefixes. Each strain name prefix is used by only one laboratory (Appendix A) and should be registered with the central clearing house. Strain name prefixes should not end in "i" or "l" or "o". For example, CB61 and MT177 are strain names.

Occasionally, strains may be generated that carry the same set of known mutations as strains that already exist. Such strains must be given a new strain name. Mutations are never renamed.

VI. Hybrid Strains

Alleles may be introduced into N2 from other strains isolated from nature, such as *C. elegans* Bergerac (e.g., Fatt and Dougherty, 1963). Such alleles and any new genes they define should be named according to the system of nomenclature described above.

VII. Phenotypic Abbreviations

Phenotypic characteristics generally should be described in words, e.g. "uncoordinated progeny were picked". If more convenient, a non-italicized abbreviation, which can correspond to a gene name, may be used; the first letter of a phenotypic abbreviation is capitalized. For example, "Dpy animals segregated both Dpy Unc and Dpy Tra progeny". Also, "Heterozygotes segregated Dpy Unc-52 and Unc-4 progeny." Wild-type phenotypes can be indicated using the prefix "non", as in "Dpy non-Unc and Unc non-Dpy recombinants were selected". Phenotypic abbreviations that do not correspond to gene names should be defined explicitly when first introduced.

Lower case italicized gene names should not be used to designate phenotypic traits.

In certain instances, it may be necessary to refer to a mutation for which genetic analysis is incomplete. Generally, the mutation name alone should be used. If convenient for discussion purposes, an appropriate phenotypic or genotypic abbreviation may be included, e.g. Unc*(e1500)* or *lin(e1413)*.

VIII. Revisions of Existing Names

Gene, mutation, and strain names that have been published and that conform to the above system should not be altered. Published names that do not conform should be changed appropriately and the change noted when the new designation is first published.

Acknowledgements. This system of nomenclature has been established in collaboration with many colleagues, most of whom met at Woods Hole, Massachusetts, in April, 1977, and at Cold Spring Harbor, New York, in May, 1979.

References

Barratt, R.W., Perkins, D.D.: *Neurospora* genetic nomenclature. Neurosopora Newsletter **8**, 23–24 (1965)

Borstel, R.C. von (ed.): Proposal for a new system of genetic nomenclature for use in yeast genetics research. Yeast genetics supplement to Microbial Genetics Bulletin, No. 31, pp. 1–8 (1969)

Brenner, S.: The genetics of *Caenorhabditis elegans*. Genetics **77**, 71–94 (1974)

Committee on Standardized Genetic Nomenclature for Mice. A revision of the standardized genetic nomenclature for mice. J. Hered. **54**, 159–162 (1963)

Clutterbuck, A.J.: Gene symbols in *Aspergillus nidulans*. Genet. Res. **21**, 291–296 (1973)

Demerec, M., Adelberg, E.A., Clark, A.J., Hartman, P.E.: A proposal for a uniform nomenclature in bacterial genetics. Genetics **54**, 61–76 (1966)

Fatt, H.V., Dougherty, E.C.: Genetic control of differential heat tolerance in two strains of the nematode *Caenorhabditis elegans*. Science **141**, 266–267 (1963)

Herman, R.K.: Crossover suppressors and balanced recessive lethals in *Caenorhabditis elegans*. Genetics **88**, 49–65 (1978)

Herman, R.K., Albertson, D.G., Brenner, S.: Chromosome rearrangements in *Caenorhabditis elegans*. Genetics **83**, 91–105 (1976)

Lindsley, D.L., Grell, E.H.: Genetic variations of *Drosophila melanogaster*. Carnegie Institution of Washington Publication No. 627, Oak Ridge, Tennessee (1968)

Communicated by L. Lerman

Received May 30, 1979

Appendix A

Laboratory abbreviations for mutation and strain names

Letter prefixes		Laboratory representative and current address
Mutation	Strain	
a	GT	D. Dusenbery, School of Biology, Georgia Institute of Technology, Atlanta, GA
b	DH	D. Hirsh, Department of Molecular, Cellular, and Developmental Biology, University of Colorado, Boulder, CO
bb	BM	D. Mitchell, Boston Biomedical Research Institute, 20 Staniford St., Boston, MA
ct	BW	W. Wood, Department of Molecular, Cellular, and Developmental Biology, University of Colorado, Boulder, CO
e	CB	J. Sulston, MRC Laboratory of Molecular Biology, Cambridge, England
f	FF	J. Brun, Department of Biologie Generale et Appliqué, Universite Claude Bernard Lyon-I, 43 Bd. du 11 Novembre 1918, 69621 Villeurbanne, France
g	GG	G. von Ehrenstein, Max Planck Institut für Experimentelle Medizin, 3400 Göttingen, Hermann Rein Straße 3, West Germany
hc	BA	S. Ward, Department of Embryology, Carnegie Institution of Washington, Baltimore, MD
hs	HH	R. Hecht, Department of Biophysical Sciences, University of Houston, Houston, TX
m	DR	D. Riddle, Division of Biological Sciences, University of Missouri, Columbia, MO
mn	SP	R. Herman, Department of Genetics and Cell Biology, University of Minnesota, St. Paul, MN
n	MT	R. Horvitz, Department of Biology, Massachusetts Institute of Technology, Cambridge, MA
p	PR	R. Russell, Department of Life Sciences, University of Pittsburgh, Pittsburgh, PA
q	JK	J. Kirschbaum, Children's Hospital Medical Center, Department of Neuroscience, Boston, MA

Letter prefixes		Laboratory representative and current address
Mutation	Strain	
rp	NE	R. Pertel, US EPA, CMZ, Room 1006 WH568, Metabolic Effects Branch, 1921 Jefferson Davis Highway, Arlington, VA
s	BC	D. Baillie, Department of Biological Sciences, Simon Fraser University, Burnaby, British Columbia, Canada
sc	BE	R. Edgar, Thimann Laboratory, University of California, Santa Cruz, CA
st	RW	R. Waterston, Department of Anatomy and Neurobiology, Washington University of St. Louis, MO
su	HE	H. Epstein, Department of Neurology, Baylor College of Medicine, Texas Medical Center, Houston, TX
sz	AF	A. Fodor, Biological Research Center, Department of Genetics, Hungarian Academy of Sciences, P.O. Box 521, H-6701 Szeged, Hungary
t	TT	P. Babu, Tata Institute of Fundamental Research, Bombay, India
w	WW	M. Samoiloff, Department of Zoology, University of Manitoba, Winnepeg, Canada
x	ZZ	J. Lewis, Department of Biological Sciences, Columbia University, New York, NY

Appendix B

Column I: general gene name
Column II: laboratory assigning gene name (listed by laboratory mutation name prefix as in Appendix A)
Column III: phenotype

I	II	III
ace	*p*	acetylcholinesterase abnormality
bli	*e*	blistered
cat	*e*	abnormal catecholamine
che	*e*	abnormal chemotaxis
daf	*e*	abnormal dauer formation
dpy	*e*	dumpy
emb	*g*	abnormal embryogenesis
fer	*hc*	abnormal fertilization
flu	*e*	abnormal fluorescence
her	*e*	hermaphroditization
him	*e*	high incidence males
isx	*hc*	intersex
lan	*e*	abnormal lannate sensitivity
let	*mn*	lethal
lev	*x*	abnormal levamisole sensitivity
lin	*e*	abnormal lineage
lon	*e*	long
mab	*e*	male abnormal
mec	*e*	mechanosensory abnormality
nuc	*e*	abnormal nuclease
osm	*p*	abnormal osmotic avoidance
rol	*e*	roller
sma	*e*	small
sns	*p*	sensory abnormality
sqt	*sc*	squat
sup	*e*	suppressor
tax	*a*	abnormal chemotaxis
ttx	*p*	abnormal thermotaxis
tra	*e*	transformer
unc	*e*	uncoordinated
vab	*e*	variable abnormal