GENETIC NOMENCLATURE

Strain collections. The ease of rapidly accumulating a large number of mutants requires careful bookkeeping to avoid confusing one mutant with another. Each mutant should be assigned a strain number. Strain numbers usually consist of 2-3 capital letters designating their source and a serial numbering of the strains in a central laboratory collection. It is a good idea to check existing genetic resources to avoid the potential confusion that can result from assigning different genes the same name. In addition to genome databases, good resources for gene names include the *Salmonella* Genetic Stock Centre (http://www.ucalgary.ca/~kesander/), the *E. coli* Genetic Stock Center (http://cgsc.biology.yale.edu/), and the *Bacillus subtilis* Genetics web site (http://www1.rhbnc.ac.uk/biological-sciences/cutting/index.html).

Nomenclature. Through the 1960's, genetic nomenclature was a "tower of babel". Due to the absence of clear rules for naming genes, each investigator assigned new names haphazardly, often resulting in the same name being applied to different genes or different names being applied to the same gene. To eliminate the resulting confusion, Demerec et al. (1966, 1986) developed a standard nomenclature for bacterial genes. With the development of new genetic tools, some modifications have been required. A detailed description of these rules can be found in the instructions to authors for the J. Bacteriol. (http://jb.asm.org/misc/ifora.shtml). The basic rules are described below.

GENOTYPE:

1. **Genes**. Each gene is assigned a three-letter designation, usually an abbreviation for the pathway or the phenotype of mutants. When the genotype is indicated, the three-letter designation is written in lower case. Different genes that affect the same pathway are distinguished by a capital letter following the three-letter designation.

   For example, mutations affecting pyrimidine biosynthesis are designated *pyr*; the *pyrC* gene encodes the enzyme dihydroorotase and the *pyrD* gene encodes the enzyme dihydroorotate dehydrogenase.

2. **Allele numbers.** Each mutation in the pathway is consecutively assigned a unique allele number. A separate series of allele numbers is used for each three-letter locus designation. If there is no capital letter designating a specific gene, insert a dash before the allele number.

   For example, *pyrC19* refers to a particular *pyr* mutation that affects the *pyrC* gene. In order to distinguish each mutation, no other *pyr* mutation, regardless of the gene affected, will be assigned the allele number *19*. A separate series of allele numbers is used for each three-letter locus designation. Allele numbers should be used sequentially and carefully monitored to insure that two different mutations are not named with the same allele numbers.

   The entire genotype is italicized or underlined (e.g. *pyrC19*).

3. **Insertions**. Transposable elements or suicide plasmids can insert in known genes or in a site on the chromosome where no gene is yet known. When an insertion is in a known gene, the mutation is given a three-letter designation, gene designation, and allele number as described above, followed by a double colon then the type of insertion element. DO NOT leave blank spaces between the letters or numbers and the colon. For example, a particular Tn*10* insertion within the *pyrC* gene (mutant allele number *103*) may be designated *pyrC103*::Tn*10*.

When a transposon insertion is not in a known gene, it is named according to the map position of the insertion on the chromosome. Such insertions are named with a three-letter symbol starting with *z*. The second and third letters indicate the approximate map position in minutes: the second letter corresponds to 10-minute intervals of the genetic map numbered clockwise from minute 0 (*a* = 0-9; *b* = 10-19; *c* = 20-29, etc.); the third letter corresponds to minutes within any 10-minute segment (*a* = 0; *b* = 1; *c* = 2; etc). For example, a Tn*10* insertion located near *pyrC* at 23 minutes is designated *zcd*::Tn*10*. Allele numbers are assigned sequentially to such insertions regardless of the letters appearing in the second and third positions, so if more refined mapping data suggests a new three-letter symbol, the allele number of the insertion mutation is retained. This nomenclature uses *zaa* (0 min) to *zjj* (99 min). Insertion mutations on extrachromosomal elements are designated with *zz*, followed by a letter denoting the element used. For example, *zzf* is used for insertion mutations on an F' plasmid. Insertions with an unknown location are designated *zxx*.

zaa =	insertion at 0-1 min		*zaa* =	insertion at 0 min
zab =	insertion at 1-2 min		*zba* =	insertion at 10 min
zac =	insertion at 2-3 min		*zca* =	insertion at 20 min
zad =	insertion at 3-4 min		*zda* =	insertion at 30 min
zae =	insertion at 4-5 min		*zea* =	insertion at 40 min
zaf =	insertion at 5-6 min		*zfa* =	insertion at 50 min
zag =	insertion at 6-7 min		*zga* =	insertion at 60 min
zah =	insertion at 7-8 min		*zha* =	insertion at 70 min
zai =	insertion at 8-9 min		*zia* =	insertion at 80 min
zaj =	insertion at 9-10 min		*zja* =	insertion at 90 min

zxx =	insertion with unknown location
zzf =	insertion on F-plasmid

Some commonly used mini-transposon derivatives are designated as follows:

Tn*10*dTet	=	Tet resistance, deleted for Tn*10* transposase
Tn*10*dCam	=	Derived from Tn10dTet, Cam resistance substituted for Tet resistance
Tn*10*dKan	=	Derived from Tn*10*dTet, Kan resistance substituted for Tet resistance
Tn*10*dGen	=	Derived from Tn*10*dTet, Gen resistance substituted for Tet resistance
MudJ	=	Kan resistance, forms *lac* operon fusions, deleted for Mu transposase
MudJ-Cam	=	Derived from MudJ, Cam resistance marker disrupts Kan resistance
MudCam	=	Cam resistance substitution between ends of Mu

4. **Plasmids.** Plasmids should be indicated by a / slash after the genotype. Indicate the name of the plasmid, the plasmid origin, and the relevant genotype or phenotype carried by the plasmid.

   Insertions of suicide plasmids into the chromosome can be indicated as described for transposons. If a duplication is generated it can be described as indicated under chromosomal rearrangements.

5. **Phage.** Prophages or plasmids integrated into an attachment site can be indicated by the name of the attachment site followed by a double colon and the phage genotype indicated in brackets. For example, *att*::[P22 *mnt*::Kan].

6. **Chromosome rearrangements.** Chromosome rearrangements including deletions, duplications, and inversions should be indicated by a three letter symbol indicating the type of rearrangement, followed by the genes involved indicated in parenthesis, followed by the allele number.

   Deletions   =   DEL(genes)allele number
   Inversions   =   INV(join point gene #1 – join point gene #2)allele number
   Duplications =   DUP(gene #1*join point*gene #2)allele number

PHENOTYPE:

1. **Growth phenotypes**. It is often necessary to distinguish the phenotype of a strain from its genotype. The phenotype is usually indicated with the same three-letter designation as the genotype but phenotypes start with capital letters and are not underlined. (For example, strain TR251 [*hisC527 cysA1349 supD*] has a Cys$^+$ His$^+$ phenotype because the s*upD* mutation suppresses the amber mutations in both the *cysA* and the *hisC* genes.)

2. **Antibiotic resistance.** Both two and three letter designations are commonly used for antibiotic resistance markers. Both are acceptable, but it is essential to be consistent. Resistance and sensitivity is indicated with a superscript but on the computer it is often simpler to indicate resistance with (R) and sensitivity with (S).

Amp	=	Ampicillin	Str	=	Streptomycin
Cam	=	Chloramphenicol	Tet	=	Tetracycline
Gen	=	Gentamicin	Zeo	=	Zeomycin
Kan	=	Kanamycin			
Neo	=	Neomycin	XG	=	X-gal
Spc	=	Spectinomycin	XP	=	X-phosphate

3. **Conditional alleles.** Conditional alleles indicated by the genotype including allele number followed by the two letter designation for the conditional phenotypes shown in parenthesis. For example, *leuA414*(Am). Note that because this is a phenotype it begins with a capital letter.

(Ts)	=	Temperature sensitive mutation	(Op)	=	Opal mutation
(Cs)	=	Cold sensitive mutation	(Oc)	=	Ochre mutation
(Am)	=	Amber mutation			

REFERENCES:

Demerec, M., E. Adelberg, A. Clark , and P. Hartman. 1966. A proposal for a uniform nomenclature in bacterial genetics. Genetics 54(1):61-76..

Demerec, M., E. Adelberg, A. Clark , and P. Hartman. 1968. A proposal for a uniform nomenclature in bacterial genetics. J Gen Microbiol. 50(1):1-14.

Maloy, S., J. Cronan, and D. Friefelder. 1994. Microbial Genetics, Second edition. Jones and Bartlett, MA.

Maloy, S., V. Stewart, and R. Taylor. 1996. Genetic Analysis of Pathogenic Bacteria. Cold Spring Harbor Laboratory Press, NY.

Journal of Bacteriology Instructions to Authors. 2001. http://jb.asm.org/misc/ifora.shtml