

## [1] Strain Collections and Genetic Nomenclature

By STANLEY R. MALOY and KELLY T. HUGHES

### Abstract

The ease of rapidly accumulating a large number of mutants requires careful bookkeeping to avoid confusing one mutant with another. Each mutant constructed should be assigned a strain number. Strain numbers usually consist of two to three capital letters designating the lab where they were constructed and a serial numbering of the strains in a central laboratory collection. Every mutation should be assigned a name that corresponds to a particular gene or phenotype, and an allele number that identifies each specific isolate. When available for a particular group of bacteria, genetic stock centers are the ultimate resources for gene names and allele numbers. Examples include the *Salmonella* Genetic Stock Centre (<http://www.ucalgary.ca/~kesander/>), and the *E. coli* Genetic Stock Center (<http://cgsc.biology.yale.edu/>). It is also important to indicate how the strain was constructed, the parental (recipient) strain, and the source of any donor DNA transferred into the recipient strain (Maloy *et al.*, 1996).

### Introduction

Through the 1960s, genetic nomenclature was a virtual “Tower of Babel.” Due to the absence of clear rules for naming genes, each investigator assigned new names based on the method of isolation, which often resulted in the same name being applied to different genes or different names being applied to the same gene. To further confuse the issues, different investigators would each assign allele numbers independently, so two different alleles might have the exact same designation. To eliminate the resulting confusion, Demerec *et al.* (1966, 1968) developed a standard nomenclature for bacterial genes. With the development of new genetic and molecular tools, some modifications have been developed to describe particular types of mutations. With the increasing ease of determining the DNA sequence of mutants, it has become commonplace to simply indicate the amino acid sequence change of an encoded protein rather than assigning an allele number. However, even when the DNA sequence of a mutation is known, a specific allele number is invaluable for tracking the history of a strain and for maintaining large strain collections. The basic rules are described next.

## Genotype

### *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 lowercase. Multiple 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. There is only one gene required for the DNA ligase function, so mutations affecting this function are simply indicated *lig*. Three-letter-only designations are also used to indicate mutations such as deletions that affect multiple genes within a multigene operon.

### *Allele Numbers*

Each mutation in the pathway is consecutively assigned a unique allele number. Even multiple mutations constructed by directed mutagenesis are assigned different allele numbers to indicate that they arose independently. A separate sequential series of allele numbers is used for each three-letter locus designation. Blocks of allele numbers are assigned to laboratories by the appropriate genetic stock center. Allele numbers should be used sequentially and carefully monitored to ensure that two different mutations are not named with the same allele numbers.

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. The entire genotype is italicized or underlined (e.g., *pyrC19*). A separate series of allele numbers is used for each three-letter locus designation. In cases where there is only a single gene in a pathway or the particular gene in the pathway is unknown, and hence there is no capital letter following the three-letter symbol, insert a dash before the allele number. For example, *lig-131* refers to a particular mutation in the *lig* gene; *pyr-67* refers to a particular mutation that disrupts the pyrimidine biosynthesis pathway, but it is not yet known which gene in this pathway is mutated.

### *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 just described, followed by a double

colon, and then the type of insertion element. *Do not* leave blank spaces between the letters or numbers and the colon. For example, a particular *Tn10* insertion within the *pyrC* gene (mutant allele number *103*) may be designated *pyrC103::Tn10*.

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 *Tn10* insertion located near *pyrC* at 23 min is designated *zcd::Tn10*. Allele numbers are assigned sequentially to such insertions regardless of the letters appearing in the second and third positions, so that 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). The map position for a given insertion might change with refined mapping resulting in letter changes (i.e., *zae* to *zaf*), but the allele number never does. It is the allele number that defines a particular mutation. Insertion mutations on extra-chromosomal 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*. Allele designation of insertion mutants in unknown genes based on chromosome map location:

<i>zaa</i>	= insertion at 0–1 min
<i>zab</i>	= insertion at 1–2 min
<i>zac</i>	= insertion at 2–3 min
<i>zad</i>	= insertion at 3–4 min
<i>zae</i>	= insertion at 4–5 min
<i>zaf</i>	= insertion at 5–6 min
<i>zag</i>	= insertion at 6–7 min
<i>zah</i>	= insertion at 7–8 min
<i>zai</i>	= insertion at 8–9 min
<i>zaj</i>	= insertion at 9–10 min
<i>zaa–zaj</i>	= insertion in 0–10 min region
<i>zba–zbj</i>	= insertion in 10–20 min region
<i>zca–zcj</i>	= insertion in 20–30 min region
<i>zda–zdj</i>	= insertion in 30–40 min region
<i>zea–zej</i>	= insertion in 40–50 min region
<i>zfa–z fj</i>	= insertion in 50–60 min region

*zga-zgj* = insertion in 60–70 min region  
*zha-zhj* = insertion in 70–80 min region  
*zia-zij* = insertion in 80–90 min region  
*zja-zjj* = insertion in 90–100 min region  
*zxx* = insertion with unknown location  
*zzf* = insertion on F-plasmid

A few commonly used minitransposon derivatives are designated as follows:

*Tn10dTet* = Tet resistance, deleted for *Tn10* transposase  
*Tn10dCam* = Derived from *Tn10dTet*, Cam resistance substituted for Tet resistance  
*Tn10dKan* = Derived from *Tn10dTet*, Kan resistance substituted for Tet resistance  
*Tn10dGen* = Derived from *Tn10dTet*, 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

### Plasmids

When writing the genotype of a strain, plasmids are often indicated by a slash (/) after the chromosome genotype. It is important to keep track of 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.

### Phages

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. An example is *att::[P22 mnt::Kan]*.

### 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 parentheses, and then the allele number (Schmid and Roth, 1983). The genes and allele

number should be italicized or underlined. Rules for this nomenclature are summarized below.

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
Unknown	= CRRallele number

## Phenotype

### *Growth Phenotypes*

It is often necessary to distinguish the phenotype of a strain from its genotype (Maloy *et al.*, 1994). The phenotype is usually indicated with the same three-letter designation as the genotype, but phenotypes start with capital letters and are not italicized or underlined. (For example, strain TR251 [*hisC527 cysA1349 supD*] has a Cys(+) His<sup>+</sup> phenotype because the *supD* mutation suppresses the amber mutations in both the *cysA* and the *hisC* genes.)

### *Antibiotic Resistance*

Both two- and three-letter designations are commonly used for antibiotic resistance markers. Both are acceptable, but be consistent. Resistance and sensitivity are indicated with a superscript. Common designations are listed below.

Ap Amp	= Ampicillin
Cm Cam	= Chloramphenicol
Gm Gen	= Gentamicin
Km Kan	= Kanamycin
Nm Neo	= Neomycin
Sp Spc	= Spectinomycin
Sm Str	= Streptomycin
Tc Tet	= Tetracycline
XG	= X-gal
XP	= X-phosphate
Zm Zeo	= Zeomycin

### *Conditional Alleles*

Conditional alleles are indicated by the genotype and allele number followed by the two-letter designation for the conditional phenotypes in parentheses. An example is *leuA414*(Am). Because the two-letter designation

is a phenotype, it begins with a capital letter. Note that Ts usually refers to heat-sensitive mutants. Common designations for conditional phenotypes follow:

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

## References

- Demerec, M., Adelberg, E., Clark, A., and Hartman, P. (1966). A proposal for a uniform nomenclature in bacterial genetics. *Genetics* **54**, 61–76.
- Demerec, M., Adelberg, E., Clark, A., and Hartman, P. (1968). A proposal for a uniform nomenclature in bacterial genetics. *J. Gen. Microbiol.* **50**, 1–14.
- Maloy, S., Cronan, J., and Friefelder, D. (1994). “Microbial Genetics,” 2nd ed. Jones and Bartlett Publishers, Boston.
- Maloy, S., Stewart, V., and Taylor, R. (1996). “Genetic Analysis of Pathogenic Bacteria.” Cold Spring Harbor Laboratory Press, NY.
- Schmid, M. B., and Roth, J. R. (1983). Genetic methods for analysis and manipulation of inversion mutations in bacteria. *Genetics* **105**, 517–537.