GOLDBIO®

Guide to *E. coli* Genotype and Genetic Marker Nomenclature

Choosing the right competent cell line for your assay and achieving a successful transformation are dependent on understanding genotypes and genetic markers. Here, we describe the nomenclature of *E. coli* genotypes to help guide you in choosing your competent cells.

Genotype

The genotype (genes) of an organism, including *E. coli*, is expressed resulting in a phenotype, which is an observable and measureable characteristic. The genotype will contain information for genes, mutation types, allele numbers, antibiotic resistance, insertions, fusions, inversions, plasmids and deletions carried by a specific bacterial strain.

Genes: In *E. coli*, the genotype only includes the genes that carry a loss of function mutation. The gene name is listed as three-letters in lower case and italics (DNA methylase is written as *dam*). Different genes affecting the same function/pathway are identified with different uppercase italic letters. *GalE* and *galK* are examples of genes encoding proteins that function in the same pathway.

- ***galE*** → Gene encoding the enzyme UDP-galactose-4-epimerase, which catalyzes conversion of UDP-galactose (UDP-Gal) to UDP-glucose (UDP-Glc). Mutation affects galactose metabolism.
- ***galK*** → Gene product is galactokinase, which catalyzes the phosphorylation of galactose to galactose-1-phosphate. Mutation results in lack of galactose metabolism.

In the case of different mutations in the same gene, or different alleles (genes can have variants, which are known as alleles), a number is added to indicate the specific mutation (allele).

- ***glnV44*** → Glutamine insertion, mutant allele number 44.

The presence of a (^{q}) indicates a constitutive mutation.

- ***lacI^q*** → Mutation (-35 site in promoter of *lacI*, GCGCAA to GTGCAA) results in constitutive expression of lac repressor and inhibition of the lac promoter. Useful when a lac operator is also present.

Insertions: Transposable elements have their own mechanism to move and insert themselves

Gold Biotechnology
St. Louis, MO
Ph: (800) 248-7609
Web: www.goldbio.com
Email: contactgoldbio86@goldbio.com

into the host organism's DNA. Insertion of a transposon into a known gene is indicated as ***gene*****::transposable element**.

- ***cysC95*****::Tn10** → Tn10 insertion within *cysC* gene, mutant allele number 95.

- **Common transposons:**
  - **Tn10** → Transposon carrying tetracycline resistance.
  - **Tn5** → Transposon carrying kanamycin resistance.

<u>**Deletions**</u>: A deletion of a specific gene is indicated with **Δ** and stated as ***Δgene****,* where Δ is placed before the gene name*.* If a deletion encompasses multiple genes, then it is written *as* ***Δ****(gene-gene)* indicating that a whole region including the genes listed and those in that segment are deleted.

- ***ΔleuA*** → Deletion of *leuA*.
- ***Δ(lac-proAB)*** → Deletion of lac operon, genes *proA* and *proB* and the genes contained in that region.

<u>**Fusions**</u>: The fusion of different elements is often indicated by **ϕ(*fused element-fused element*)**, by a colon (:) or simply by a dash (-).

- ***(gntK-lacZ)*** → Fusion of *gntK* with the *lacZ* gene.
- ***ϕ(gntK-lacZ)*** → Fusion of *gntK* with the *lacZ* gene.
- ***gntK:lacZ*** → Fusion of *gntK* with the *lacZ* gene.

<u>**Inversion**</u>: A chromosomal inversion of a segment between two specific sites is indicated by **INV(gene-gene)**. See Figure 1.

- **INV(*aroA-fadR*)** → Inversion of sequence flanked by *aroA* and *fadR*.

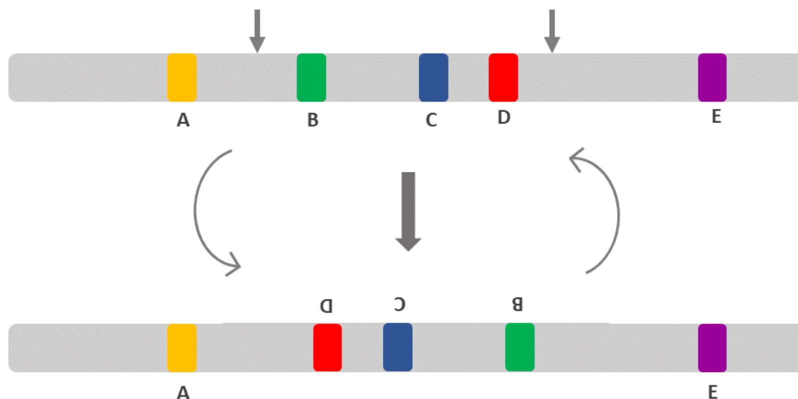

Figure 1. Inversion of a chromosomal segment containing genes B, C and D.

Plasmids: The inclusion of specific plasmids can be indicated in the genotype. The F plasmid is a DNA plasmid called Fertility Factor or Sex Factor. This plasmid carries the genes that allows the transfer of plasmid DNA from one cell to another. Confers the ability to mate with F^- through conjugation.

- **F^+** → Indicates that the bacterium carries the F plasmid. However, the F plasmid is not inserted into the bacterial genome.
- **F^-** → Indicates the bacterium does not carry F plasmid.
- **F'[ ]** → In this case, the F plasmid has inserted into the bacterial genome, then has gone through excision, becoming an independent plasmid. However, during excision, some of the bacterial genome sequences were excised with the F plasmid. Thus, this indicates that the F plasmid exists outside of the bacterial genome and carries bacterial genes, which are specified inside the brackets.
- **Hfr** → Denotes high frequency of recombination. Indicates that the F plasmid has inserted (integrated) into the bacterial genome.

Prophages: A prophage, also known as a "phage," can infect bacterial cells resulting in its genome integrating into the bacterial DNA chromosome (lysogenization) or becoming an extrachromosomal plasmid (lytic growth).

- **ϕ80** → Strain carries lambdoid prophage ϕ80. Recognizes the FhuA receptor on *E.* coli allowing binding to E. coli membrane.
- **P1** → Indicates that the strain carries the P1 prophage. Expresses the P1 restriction system.
- **P2** → Indicates that the strain carries a P2 prophage. Found in gram-negative bacteria. Useful for selection against Red+ Gam+ λ in the lambda red recombination-mediated genetic engineering system, which is a powerful cloning method.

Restriction-modification system: *E. coli* have restriction (cleavage of DNA) and modification (by methylation) systems that confer protection from foreign DNA.

The *hsdR, hsdM and hsdS* genes encode an EcoKI restriction-modification system which degrades DNA sequences that are **not** methylated at the adenine at 5'-AACN$_6$GTGC-3'. These sequences are degraded by the EcoKI restriction enzyme.

- **$r_{B/K}^{+/-}$** → r indicates the restriction system and strain is indicated by uppercase subscript B/K ($_{B/K}$). The ($^{+/-}$) indicates whether or not the strain has the restriction system.
- **$m_{B/K}^{+/-}$** → m indicates the methylation modification system and strain is indicated by

lowercase subscript B/K ($_{B/K}$). The ($^{+/-}$) indicates whether or not the strain has the modification system.

- **hsdS(r_{B^-}, m_{B^-})** → Indicates that both the restriction and methylation systems are inactive since the strain does not have an endonuclease or methylase. Thus, DNA sequences are not restricted or methylated. This genotype is useful for cloning applications since cleavage/degradation of cloned DNA does not occur.

On the other hand, McrA, McrBC and Mrr are part of different restriction-modification systems in which restriction (digestion) occurs if **DNA methylation is present**. McrA and McrBC recognize methylcytosine. Whereas, Mrr recognizes both methylcytosine and methyladenine. Mutation results in inactivation of cleaving of DNA with methylated cytosine (mutation of *mcrA*, *mcrBC*, *mrr*) or adenine (mutation of *mrr*) and allows cloning of methylated DNA.

- **Δ(*mrr-hsdRMS-mcrBC*)** → Indicates that the strain lacks *mrr*, *hsdRMS and mcrBC* genes, which is advantageous for cloning.

Other Phenotypic Traits

Relevant phenotypic characteristics, including antibiotic resistance and sensitivity to cold, can also be described in the genotype using the following specific nomenclature.

__Antibiotic resistance__: Resistance to a specific antibiotic is indicated by two, three or four letters for the antibiotic and a superscript uppercase R (^{R}).

- **Str^R, Sm^R** → streptomycin resistance
- **$Carb^R$** → carbenicillin resistance
- **Amp^R Ap^R, Amp^R** → ampicillin resistance
- **Gen^R, Gm^R** → gentamicin resistance
- **Kan^R, Km^R** → kanamycin resistance
- **Neo^R, Nm^R** → neomycin resistance
- **Zeo^R, Zm^R** → zeomycin resistance
- **Tet^R, Tc^R** → tetracycline resistance
- **Spc^R, Sp^R** → spectinomycin resistance
- **Cm^R** → chloramphenicol resistance

__Sensitivity__: Conditional alleles with different sensitivities are indicated in the genotype. Sensitivity to cold and temperature are indicated as lowercase superscript cs (^{cs}) and ts (^{ts}), respectively. Temperature sensitivity indicates that exposure to high temperature results in inactivation. Cold sensitivity results in a loss-of-function phenotype only at low temperatures (lower than a specific threshold temperature).

- ***$leuA^{ts}$*** → Temperature sensitive allele of *leuA.*

Gold Biotechnology
St. Louis, MO
Ph: (800) 248-7609
Web: www.goldbio.com
Email: contactgoldbio86@goldbio.com

- ***leuA**^{cs}* → Cold sensitive allele of *leuA.*

Other conditional alleles: Other mutations can be indicated with two lowercase superscript letters. Examples are the amber mutation, the opal mutation and the ochre mutation, which introduce a stop codon and results in a truncated protein product.

- ***argE**^{am}* → Amber mutation in *argE* introducing a nonsense codon (UAG).
- ***argE**^{op}* → Opal mutation in *argE* introducing a nonsense codon (UGA).
- ***argE**^{oc}* → Opal mutation in *argE* introducing a nonsense codon (UAA).

Relevant Genetic Markers

Gene	Description and function
ara14	Strain carries mutation resulting in inability to metabolize arabinose.
araD	This gene encodes for ribulose-5-phosphate 4-epimerase, which is one of the enzymes needed for metabolism of the sugar L-arabinose in *E. coli*. Strains with an *AraD* mutation cannot metabolize L-arabinose.
cycA	*CycA* encodes for a D-alanine, D-serine, glycine and D-cycloserine transporter. Strains carrying a *cycA* mutation cannot use D-alanine as a carbon source.
dapD	*DapD* encodes for succinyl-diaminopimelate aminotransferase. Mutation leads to requirement of succinate or a lysine and methionine mixture.
dam	Mutation in DNA adenine methyltransferase (an enzyme that methylates adenine in GATC sequence, in double stranded DNA), resulting in lack of adenine methylation, high recombination efficiency and activation of DNA repair.
dcm	Mutation in DNA cytosine methyltransferase which methylates the second C of CCWGG, preventing cleavage of DNA by restriction enzymes. Mutation is indicated with *dcm-*.
DE3	Strain carries the DE3 phage encoding for T7 RNA polymerase. Useful when inducing protein expression using T7 promoter expression systems.
deoR	This gene's product acts as transcription factor and regulates gene expression. Allows uptake of large plasmids and cloning of large sequences. Mutation results in constitutive expression of deoxyribose synthesis genes.
dut1	This mutation eliminates dUTPase (deoxyuridine triphosphatase; converts dUTP to dUMP and PPi) activity resulting in increased dUTP concentrations, leading to incorporation of uracil (instead of thymine) into DNA. Stable U incorporation also requires the *ung* gene mutation.
endA1	Mutation results in a lack of Endonuclease I. Facilitates plasmid insertion since intracellular endonucleases are not present, preventing cleavage and degradation of plasmids.

fhuA (also TonA)	Mutation in iron uptake receptor resulting in resistance to T1/T5 phage since the phage-bacteria interaction cannot occur.
galE	Mutation in UDP-glucose 4-epimerase, affects galactose metabolism, resulting in lack of growth in galactose.
galk	Mutation of galactokinase that results in inability to metabolize galactose. Confers resistance to 2-deoxygalactose.
galU	Mutation in glucose-1-phosphate uridylyltransferase leading to lack of galactose metabolism. Cannot grow in galactose only.
glnV (also supE)	Mutation in tRNA (transfer ribonucleic acid) resulting in suppression of the amber nonsense mutation (amber nonsense mutation inserts UAG stop codons leading to truncated proteins). Required for some phage growth since lower levels of translation termination are observed.
gyrA96	Mutation in DNA gyrase (an enzyme that allows ATP-dependent negative supercoiling of double-stranded circular DNA). Mutation results in nalidixic acid (a bactericidal that inhibits replication in bacteria) resistance.
hfla150	Mutation of protease HflKC–a membrane protein that degrades the cII protein (a protein important for the life cycle of λ phage). Thus, phage cII proteins are stabilized and a high frequency of lysogenization by λ results.
lacZ	This gene encodes β-galactosidase and is part of the *lac* operon. β-galactosidase cleaves lactose.
lacZΔM15	Partial deletion of the *lacZ* gene in the alpha region allowing α complementation. Used in blue/white selection on XGal plates.
lacI^q	Mutation results in constitutive expression of the *lac* repressor protein.
lacY	Lactose permease mutation resulting in deficient lactose transport.
lon	Lon is a cytoplasmic protease that degrades aberrant proteins in bacteria in response to stress. Δ*lon* results in lack of proteolysis and may facilitate expression of a protein of interest. However, Lon may be useful in membrane protein overexpression. B strains do not express Lon protease.
mcrA, mcrBC, mrr	Mutation results in inactivation of cleaving of DNA with methylated cytosine (*mcrA*, *mcrBC*, *mrr*) or adenine (*mrr*) and allows cloning of methylated DNA.
ompT	Mutation in outer membrane protein protease VII resulting in lower proteolysis of expressed proteins.
pLysS	Indicates that the strain carries the *pLysS* plasmid, which confers chloramphenicol resistance and produces T7 lysozyme. T7 lysozyme is an effective inhibitor of T7 RNA polymerase. Used for better inhibition of expression under non-induced conditions.

Gold Biotechnology
St. Louis, MO
Ph: (800) 248-7609
Web: www.goldbio.com
Email: contactgoldbio86@goldbio.com

proAB	Mutation in proline metabolism genes A and B resulting in a requirement for proline in the medium to grow.
recA1	Mutation resulting in RecA inactivation. Wildtype RecA protein mediates DNA repair and homologous recombination. Thus, the *recA1* mutation results in reduced recombination of plasmid with host DNA. The *recA1* mutation is advantageous in applications where recombination is not wanted. For example, strains carrying *recA1* are preferred for cloning since it confers cloned DNA stability.
relA1	Mutation in guanosine 3',5'-bispyrophosphate (ppGpp) synthetase I resulting in lack of ppGpp, which is a molecule normally needed to signal for a bacterial stringent response during specific stress, including amino acid starvation. In this response, RNA synthesis is inhibited. Thus, in *relA1*, RNA synthesis does occur during amino acid starvation.
Tn10	Transposon conferring tetracycline resistance.
Tn5	Transposon conferring kanamycin resistance.
λ	Lambda phage infects bacteria and is routinely used as a vector in cloning. λ^{-} denotes lambda phage deletion.
ΔH1	Deletes cro and all genes to the right of it. The Cro proteins are transcription factors that promote phage growth.

References

Casali, N., & Preston, A. (2003). *E. coli Plasmid Vectors: Methods and Applications*. Totowa, NJ: Humana Press Inc.

E. coli Genetic Resources at Yale CGSC, The Coli Genetic Stock Center. (n.d.). Retrieved November 13, 2019, from https://cgsc.biology.yale.edu/.

E. coli genotypes. (n.d.). Retrieved November 13, 2019, from https://openwetware.org/wiki/E._coli_genotypes#References.

Kihara, A., Akiyama, Y., & Ito, K. (2001). Revisiting the Lysogenization Control of Bacteriophage λ. *Journal of Biological Chemistry*, *276*(17), 13695–13700. doi: 10.1074/jbc.m011699200.

Marinus, M. G. (2000). Recombination Is Essential for Viability of an *Escherichia coli* dam (DNA Adenine Methyltransferase) Mutant. *Journal of Bacteriology*, *182*(2), 463–468. doi: 10.1128/jb.182.2.463-468.2000.

Murphy, K. C. (2016). λ Recombination and Recombineering. *EcoSal Plus*, *7*(1). doi: 10.1128/ecosalplus.esp-0011-2015.

Raleigh, E. A., Benner, J., Bloom, F., Braymer, H. D., Decruz, E., Dharmalingam, K., … Kretz, P. L. (1991). Nomenclature relating to restriction of modified DNA in *Escherichia coli*. *Journal of Bacteriology*, *173*(8), 2707–2709. doi: 10.1128/jb.173.8.2707-2709.1991.

Samuelson, J. C. (2010). Recent Developments in Difficult Protein Expression: A Guide to *E. coli* Strains, Promoters, and Relevant Host Mutations. *Methods in Molecular Biology Heterologous Gene Expression in E.coli*, 195–209. doi: 10.1007/978-1-61737-967-3_11.

Trun, N. J., Trempy, J. E., & Trempy, J. (2009). *Fundamental Bacterial Genetics*.