Denotation of *E.coli* Genotypes

A genotype indicates the genetic state of the DNA in an organism, allowing scientists to imagine how it looks like (phenotype). A genotype shows only the mutations present in *E.coli* genome. Non-indicated genes do not contain mutations. However, when using *E.coli* strains in laboratories for a long time, unidentified mutations can be accumulated. Generally these mutations may not affect results obtained by the users, but sometimes cause serious experimental problems. Therefore, it is essential to check the authenticity of strains occasionally.

***E.coli* strains :** Almost all *E.coli* strains used in laboratories have been derived from the K-12 strain. The name of K-12 strain is mentioned only if there is a difference. For example, a genotype of *E.coli* derived from the B strain should be mentioned along with its origin of strain. The original B strain genotype is londcm-. *E.coli* strains derived from K-12 are noted only when they lack a prophage or plasmid (e14, rac). However, K-12 should be indicated when bacteriophage λ is present in cells. If they harbor an F' factor and its derivatives are present, the strain should be indicated every time. The presence of a prophage or a plasmid is indicated in a bracket.

Name of genes : Genes are named in three italic lowercase letters according to their functions. For example, the dam gene comes from the first three letters of **D**NA **a**denine **m**ethylase. Many genes related to their function can be distinguished by a different capital letter at their end. For example, recA, recB, recC, and recD are different genes involved in **rec**ombination.

Dominant and recessive genes : A gene with neither + nor - as a superscript refers to a recessive gene. Sometimes, the superscript of + or - is indicated in order to avoid confusion; in a superscript letter, + refers to wild type or dominant genes, whereas - refers to mutant or recessive genes. For example, F'lac-proA+B+ contains the lac, proA, and proB genes in an F' plasmid The lac - gene is a recessive mutant gene, whereas proA and proB genes are wild type.

Deletion mutation : The greek letter Δ refers to deletion of a gene. For example, Δ(lac-pro) refers to the deletion of all the genes from lac to pro gene in the *E.coli* chromosome.

Alleles : Numbers in italic are used to describe a single gene with different mutations (collectively called alleles) For example, hsdR2, hsdR4, and hsdR17 all have different mutations in the same hsdR gene. Thus, they are referred to as alleles of hsdR. Nonsense mutations (stop codon mutations within an open reading frame) and temperature-sensitive mutations are given with specific symbols. For example, an amber mutation (a change of amino acid codon into UAG stop codon) is denoted as "am," while a mutation which causes loss of function at high temperatures as "ts" (temperature-sensitive).

Phenotype : The phenotypes of a strain are any observable characteristics which are usually expression of underlying genotypes. For example, the phenotype of Lac- means that cells with this phenotype fail to grow on lactose as a sole carbon source. The phenotype is indicated by capitalizing the first letter of genotype, always followed by superscripted + (wild type) or - (mutant), for (resistant) or s (sensitive). If a phenotype is difficult to judge from its genotype, a phenotype can be indicated in a parenthesis. For example, rpsL104(Strr) is derived from **r**ibosomal **p**rotein **s**mall subunit, S12, the mutation of which confers resistance to streptomycin as indicated in a parenthesis.

Restriction and modification : *E.coli* has a defense mechanism that identifies and destroys invading foreign DNA (restriction function). Its own DNA is modified by transferring a methyl group on a specific base on its own DNA, thereby protecting its own DNA from being destroyed (modification function) unlike the foreign DNA. Most of the *E.coli* strains process dam, dcm, and EcoK I methylation for modification functions. However, most of the K-12 and B strains for laboratory use lack EcoR I restriction /modification system. *E.coli* strains used for cloning purposes contain various mutations related to restriction and modification. To avoid any confusion, the phenotype of this mutant is indicated after a gene. The letters r and m indicate restriction and modification, respectively. The subscript denotes a strain. For example, in hsdR (rK-mK +), a mutant form of the hsdR gene, "r-" refers to loss of its restriction function, while "m+" refers to the fact that its modification function is intact. The subscript K refers to the name of K-12 strain.

Insertion : When a mobile genetic element (referred to as transposon or insertional element) or a genetic marker is inserted into a gene, a double colon (::) is placed between the two genes; one is a target gene, and the other is an inserted one. Transposons and insertional elements are abbreviated a "Tn" and "IS," respectively. For example, tetracycline resistance (Tetr) occurs because the tet gene of Tn10 transposon is inserted into the thr gene of E. coli. Thus, this insertion is indicated as thr::Tn10 (Tetr).

Plasmid : Lowercase "p" should be placed in front of a plasmid name. For example, pACYC184. An F factor does not have the lower case "p." Genotypes of an F factor can be distinguished from those of chromosome by placing "/" between them.

Distinction of genes and proteins : Generally, genes are denoted as italic lowercase letters, proteins as Roman letters with the first letter in uppercase. For example, the 'malE' gene codes for the 'MalE' protein, a maltose binding protein.

Inversion : Genes inverted are indicated with "IN" in front of the name of genes. Inversion causes genes or operon simply to run in an opposite direction in chromosome with respect to wild type, and thus most of the time phenotypes usually remain the same as wild type. For example, IN(rrnE-rrnE) - the rrnE operon is inverted with respect to wild type one.

Antibiotics : Antibiotics are indicated with following abbreviation. Amp, ampicillin; Cam, chloramphenicol; Kan, kanamycin; Nal, nalidixic acid; Str, streptomycin; Tet, tetracycline Important information on *E. coli* genes is available in an article by Berlyn, M. (1996) in Escherichia coli and Salmonella: Cellular and Molecular Biology, ed. Niehardt et al., ASM Press)

E.Coli Strain Genotype

Strain	Genotype
BL21 Star™(DE3)	F- ompT hsdSB (rB-, mB-) gal dcm rne131 (DE3)
BL21 Star™(DE3)pLysS	F- ompT hsdSB (rB-, mB-) gal dcm rne131 (DE3) pLysS (CamR)
BL21(DE3)	F- ompT hsdSB (rB-, mB-) gal dcm (DE3)
BL21(DE3)pLysE	F- ompT hsdSB (rB-, mB-) gal dcm (DE3) pLysE (CamR)
BL21(DE3)pLysS	F- ompT hsdSB (rB-, mB-) gal dcm (DE3) pLysS (CamR)
BL21-AI™	F- ompT hsdSB(rB-, mB-) gal dcm araB::T7 RNAP-tetA
BL21-SI™	F- ompT hsdSB(rB-, mB-) gal dcm endA1 lon- proUp::T7 RNAP::malQ-lacZ (TetS)
DB3.1™	F- gyrA462 endA Δ(sr1-recA) mcrB mrr hsdS20 (rB-, mB-) supE44 ara14 galK2 lacY1 proA2 rpsL20(Smr) xyl5 Δleu mtl1
DH10Bac™	F- mcrA Δ(mrr-hsdRMS-mcrBC) φ80lacZΔM15 ΔlacX74 endA1 recA1 Δ(ara, leu)7697 araD139 galU galK, nupG rpsL λ-
DH10B™	F- mcrA Δ(mrr-hsdRMS-mcrBC) φ80lacZΔM15 ΔlacX74 recA1 araD139 Δ(ara-leu)7697 galK rpsL(StrR) endA1 nupG
DH10B™-T1R	F- mcrA Δ(mrr-hsdRMS-mcrBC) φ80lacZΔM15 ΔlacX74 recA1 endA1 araD139 Δ(ara, leu)7697 galU galK λ- rpsL nupG tonA
DH12S™	F′ {proAB+ lacIqZΔM15 Tn10(TetR)} φ80lacZΔM15 recA1 Δ(mcr-hsdRMS-mcrBC) ΔlacX74 Δ(ara-leu)7697 araD139 galU galK nupG rpsL relA1
DH5α-E™	F- φ80lacZΔM15 Δ(lacZYA-argF) U169 endA1 recA1 hsdR17 (rk-, mk+) thi-1 phoA supE44 λ- gyrA96 relA1 gal-
DH5αF′IQ™	F′ proAB+ lacIqZΔM15 zzf::Tn5 (KmR) φ80lacZΔM15 Δ(lacZYA-argF) U169 recA1 endA1 hsdR17 (rk-, mk+) phoA supE44 λ-thi-1 gyr96 relA1
DH5α-FT™	F′ proAB+ lacIqZΔM15 Tn10(TetR) φ80lacZ ΔM15 Δ(lacZYA-argF) U169 recA1 endA1 hsdR17 (rk-, mk+) phoA supE44 λ-thi-1gyrA96 relA1
DH5αMCR™	F- mcrA Δ(mcr-hsdRMS-mcrBC) φ80lacZM15 (lacZYA-argF) U169 endA1 recA1 supE44 thi-1 gyrA96 relA1
DH5α™	F- φ80lacZΔM15 Δ(lacZYA-argF) U169 endA1 recA1 hsdR17 (rk-, mk+) supE44 thi-1 gyrA96 relA1 phoA
GeneHogs®	F- mcrA Δ(mrr-hsdRMS-mcrBC) φ80lacZΔM15 ΔlacX74 recA1 araD139 Δ(ara-leu)7697 galU galK rpsL (StrR) endA1 nupG fhuA::IS2
GI698 and GI724	F - λ- lacIq lacPL8 ampC::Ptrp cI Note: GI698 has no ribosome binding site before the cI repressor gene, causing decreased production of cI repressor, when compared to GI724.
INV110	F′ {traD36 proAB+ lacIq lacZΔM15} rpsL (StrR) thr leu endA thi-1 lacY galK galT ara tonA tsx dam dcm supE44 Δ(lac-proAB) Δ(mcrC-mrr)102::Tn10 (TetR)
INVαF′	F′ endA1 recA1 hsdR17 (rk-, mk+) supE44 thi-1 gyrA96 relA1 φ80lacZΔM15 Δ(lacZYA-argF) U169
LMG194	F- ΔlacX74 galE thi rpsL ΔphoA (Pvu II) Δara714 leu::Tn10 (TetR)
Mach1™-T1R	F′ φ80(lacZ)ΔM15 ΔlacX74 hsdR(rk-, mk+) ΔrecA1398 endA1 tonA
MC1061/P3	F- hsdR (rk-, mk+) araD139 Δ(araABC-leu)7679 galU galK ΔlacX74 rpsL (StrR) thi mcrB {P3: KanR AmpR (am) TetR (am)}
OmniMAX™-T1R	F- mcrA Δ(mrr-hsdRMS-mcrBC) φ80(lacZ)ΔM15 Δ(lacZYA-argF) U169 endA1 recA1 supE44 thi-1 gyrA96 relA1 tonA panD/F' proAB+ lacIq lacZΔM15 Tn10 (TetR)
PIR1	F- Δlac169 rpoS(am) robA1 creC510 hsdR514 endA recA1 uidA(ΔMlu I)::pir-116
PIR2	F- Δlac169 rpoS(am) robA1 creC510 hsdR514 endA recA1 uidA(ΔMlu I)::pir
Stbl2™	F- mcrA Δ(mcrBC-hsdRMS-mrr) recA1 endA1 lon gyrA96 thi-1 supE44 relA1 λ- Δ(lac-proAB)
Stbl4™	F′ proAB+ lacIqZΔM15 Tn10(TetR) λ- mcrA Δ(mcrBC-hsdRMS-mrr) recA1 endA1 gal supE44 gyrA96 thi-1 relA1 Δ(lac-proAB)
TOP10	F- mcrA Δ(mrr-hsdRMS-mcrBC) φ80lacZΔM15 ΔlacX74 recA1 araD139 Δ(ara-leu)7697 galU galK rpsL (StrR) endA1 nupG
TOP10/P3	F- mcrA Δ(mrr-hsdRMS-mcrBC) φ80lacZΔ {P3: KanR AmpR (am) TetR (am)}
TOP10F′	F′{lacIq Tn10(TetR)} mcrA Δ(mrr-hsdRMS-mcrBC) φ80lacZΔM15 ΔlacX74 recA1 araD139 Δ(ara-leu)7697 galU galK rpsL (StrR) endA1 nupG

Genetic Marker

Symbol	Description
ara-14	Mutation in arabinose metabolism. Blocks arabinose catabolism
argF	Ornithine carbamoyltransferase mutation blocks ability to use arginine
dam/dcm	Abolishes endogenous adenine methylation at GATC sequences (dam) or cytosine methylation at CCWGG sequences (dcm). Used to propagate DNA for cleavage with certain restriction enzymes (e.g. Ava II, Bcl I)
DE3	Lysogen that encodes T7 RNA polymerase. Used to induce expression in T7-driven expression systems
endA	endA Mutation in the non-specific endonuclease Endonuclease I; eliminates non-specific endonuclease activity, resulting in improved plasmid preps
F′	A self-transmissible, low-copy plasmid used for the generation of single-stranded DNA when infected with M13 phage; may contain a resistance marker to allow maintenance and will often carry the lacI and lacZΔM15 genotypes
galK	Galactokinase mutation blocks catabolism of galactose—cells that are galK minus grow in the presence of galactose as the sole carbon source
galU	Glucose-1-phosphate uridylyltransferase mutation blocks ability to use galactose—cells that are galU minus can grow on media that contains galactose as the sole carbon source
gyrA96	DNA gyrase mutant produces resistance to nalidixic acid
hsd	Mutations in the system of methylation and restriction that allow E. coli to recognize DNA as foreign. The hsd genotype allows efficient transformation of DNA generated from PCR reactions *hsdR–eliminates restriction of unmethylated EcoK I sites. (1) **hs
lacI	Encodes the lac repressor that controls expression from promoters that carry the lac operator; IPTG binds the lac repressor and derepresses the promoter; often used when performing blue/white screening or to control expression of recombinant genes
lacY1	Blocks use of lactose via β-D-galactosidase mutant
lacZ	β-D-galactosidase gene; mutations yield colorless (vs. blue) colonies in the presence of X-gal
lacZΔM15	Element required for β-galactosidase complementation when plated on X-gal; used in blue/white screening of recombinants; usually carried on the lambdoid prophage φ80 or F′
leuB	Requires leucine for growth on minimal media via β-isopropyl malate dehydrogenase mutation
lon	lon Deficiency in the Lon ATPase-dependent protease; decreases the degradation of recombinant proteins; all B strains carry this mutation
nupG	Mutation for the transport of nucleosides
ompT	Indicates that the E. coli lack an outer membrane protease—reduces degradation of heterologous the strains and recovery of intact recombinant proteins is improved in ompT minus strains
P3	A 60-kb low-copy plasmid that carries the ampicillin and tetracycline resistance genes with amber mutations; used predominantly for selection of supF-containing plasmids; carries the kanamycin resistance gene for selection
pLys	pLys Plasmid that encodes T7 lysozyme; used to reduce basal expression in T7-driven expression systems by inhibiting basal levels of T7 RNA polymerase
proAB	proAB Requires proline for growth on minimal media
recA	Mutation in a gene responsible for general recombination of DNA; particularly desirable when cloning genes with direct repeats
relA	RNA is synthesized in absence of protein synthesis (relaxed phenotype) relA locus regulates the coupling between transcription and translation. In the wild type, limiting amino acid concentrations results in the shutdown of RNA synthesis (also known as th
rpsL	Confers resistance to streptomycin (this makes a mutant ribosomal protein, small subunit, the target of the drug)
supE,F	tRNA glutamine suppressor of amber (supE)(UAG) or tyrosine (supF)
thi-1	Requires thiamine for growth on minimal media
Tn10	Confers tetracycline resistance via a transposon
tonA	Confers resistance to the lytic bacteriophage T1, T5 and f80
traD, D36	Prevents transfer of F' episome via transfer factor mutation
tsx	Confers resistance to phage T6 and colicin K
xyl-5	Blocks catabolism of xylose