are commonly carried on F′ factors.

The F factor is found in three alternative forms: as double-stranded, single-copy, circular extrachromosomal plasmid DNA (F^+); as plasmid DNA like F^+ but also including other bacterial genes (F′); and as a stretch of linear DNA integrated into various sites on the bacterial chromosome (Hfr). Possession of the F factor confers on *E. coli* the ability to donate DNA in bacterial crosses (or matings). For this reason cells that carry F are sometimes called *male*. F or F′ plasmids can transfer themselves to other cells, and may occasionally cause transfer of other plasmids. This latter process is called *mobilization*. Mutations called *tra* prevent F from transferring itself or mobilizing other plasmids. Integrated Tra^+ F factors (Hfr) can cause transfer of chromosomal DNA to other cells, but the recipient usually does not receive the F sequence.

NONSENSE SUPPRESSORS

Some vectors used in recombinant DNA research (e.g., plasmid πVX and phage Charon 4a) contain nonsense mutations in essential genes. These vectors must be propagated in special strains of *E. coli*. In these strains, translation of messages does not always stop when the ribosome encounters a chain termination codon (amber or ochre), but sometimes continues with a new amino acid installed at the end of the growing polypeptide. This process is called *nonsense suppression* and strains of *E. coli* in which it occurs are said to contain *nonsense suppressors*. In a strain that contains an efficient or a "strong" suppressor, suppression might occur 50% of the time an amber codon is encountered.

The mechanism of nonsense suppression is a simple one: the cell contains a mutant species of tRNA in which the anticodon loop has mutated so that it base pairs with the UAG amber codon or the UAA ochre codon. Nonsense suppressors commonly used in cloning technology are given in Table 1.4.3. UGA (opal) suppressors also exist but are rarely used.

GENETIC MARKERS

Genetic markers in *E. coli* are named according to the convention proposed by Demerec et al. (1966). All genes of a given strain are presumed to be in the wild-type state unless otherwise noted in the genotype (see box). Gene names have three italicized lowercase letters, sometimes followed by an italic uppercase letter, and sometimes also followed by an italic arabic number that specifies the precise mutation (allele) in question (e.g., *lacY1*, *trp-31*). The three-letter combination is usually a mnemonic intended to suggest the function of the gene. Proper notation omits superscript + and − in a genotype, but these are sometimes used redundantly for clarity. Deletion mutations are described by Δ, followed by the names of deleted genes in parentheses, followed by the allele number [e.g., Δ(*lac-pro*)X111]. The delta may be replaced by "del" or "d." Sometimes a phenotype designation (see box) in parentheses follows the genotype designation, if the former is not obvious from the latter [e.g., *rpsL104* (Str^r)]. However, this usage is by no means universal.

Table 1.4.4 lists commonly used genetic markers, with methods for verifying their presence or absence in bacterial cells. Genotypes of several strains used for different applications are listed in Table 1.4.5.

Table 1.4.3 Commonly Used Nonsense Suppressors[a]

Suppressor	Map position[b]	Type of suppressor	Amino acid inserted	tRNA gene
supD (su1)	43	Amber	Serine	*serU*
supE (su2)	16	Amber	Glutamine	*glnU*
supF (su3)	27	Amber	Tyrosine	*tyrT*
supB (suB)	16	Ochre/amber	Glutamine	*glnU*
supC (suC)	27	Ochre/amber	Tyrosine	*tyrT*

[a]Data compiled from Bachmann (1983) and Celis and Smith (1979).
[b]Given in minutes; see Bachmann (1983) for description.

GENOTYPE AND PHENOTYPE

Genotype indicates what genes are mutated in a strain. A genotype is a theoretical construct describing a genetic constitution that would explain the phenotype of the strain. It is derived from considerations of the strain's behavior and ancestry.

Phenotype describes the observable behavior of the strain—e.g., Lac⁻ fails to grow on lactose as a sole carbon source. Phenotypes are in Roman type, the first letter is capitalized, and the letters are always followed by superscript + or − (sometimes r, resistant, or s, sensitive). A phenotype is a datum to be explained.

Genotype and phenotype names are usually related, but the relationship is not always obvious. Examples are provided below.

Genotype	Phenotype	Description of phenotype
Some straightforward examples:		
trp-31	Trp⁻	Requires tryptophan for growth on minimal media
uvrA	UVˢ	Sensitive to UV light
recA	Rec⁻	Recombination defective
Some common examples that are not so straightforward:		
supE44	Sup⁺	Carries a tRNA suppressor gene. The *mutant* gene product, not the wild type, suppresses nonsense mutations; wild type is indicated as *sup°*, Sup⁻, and does not suppress
rpsL104	Strʳ	Resistant to streptomycin (this makes a mutant *r*ibosomal *p*rotein, *s*mall subunit, the target of the drug)
rpsE	Spcʳ	Resistant to spectinomycin (also codes for a ribosomal protein, a different one)
gyrA	Nalʳ	Resistant to nalidixic acid (the mutation affects DNA gyrase)
One mutation may create several phenotypes:		
dam-3	Dam⁻	DNA not methylated at adenines in GATC
	2-APˢ	Sensitive to 2-aminopurine
	UVˢ	Sensitive to UV light
hsdS	*Eco*K R⁻, *Eco*K M⁻	Neither restricts nor modifies DNA that enters the cell
Some mutations lead to counterintuitive phenotypes:		
recD	ExoV⁻ but Rec⁺	Exonucleolytic activity of the RecBCD protein is defective, but the recombinational activity is intact

DNA RESTRICTION, MODIFICATION, AND METHYLATION

This section and the next two describe *E. coli* functions that can prevent cloning the sequence of interest. *E. coli* has at least four restriction systems that identify foreign DNA and destroy it. These systems, encoded by *hsdRMS*, *mcrA*, *mcrB*, and *mrr*, can be avoided by using host strains in which they are disabled by mutation. Restriction of DNA and the content of methylated bases in the DNA are interrelated as described below. To select the appropriate strain, it is necessary to know the content of methylated bases in the DNA to be cloned.

The *Eco*K restriction system, encoded by the *hsdRMS* genes, is the best understood of

Table 1.4.4 Commonly Used Genetic Markers and How to Test Them[a]

Marker	Test
Nutritional markers	Streak or replica plate colonies of the strain onto plates with and without the nutrient to be tested, but which contain all other necessary nutrients.
Antibiotic resistance markers	Streak or replica plate colonies of the strain onto plates with and without the antibiotic.
Other markers	
lacZ$^+$	Streak strain on an LB plate with Xgal and IPTG (UNIT 1.4). Colonies should turn blue. Colonies of control *lacZ*$^-$ *strain should not turn blue.*
lacZΔM15[b]	Transform strain with pUC plasmid and with control plasmid such as pBR322. Streak transformants onto LB/ampicillin plate with Xgal and IPTG. Colonies bearing pUC plasmid should turn blue, while colonies bearing pBR322 should not.
F$^+$ or F′	Spot M13 phage onto a lawn of the cells. Small plaques should appear (see UNIT 1.15).
recA	Using a toothpick, make a horizontal stripe of cells across an LB plate. Also make a stripe of *recA*$^+$ control cells. Cover half of the plate with a piece of cardboard, and irradiate the plate with 300 ergs/cm^2 of 254 nm UV light from a hand-held UV source (typically 20 sec exposure from a lamp held 50 cm over the plate). *recA*$^-$ cells are very sensitive to killing by UV light, and *recA*$^-$ cells in the unshielded part of the plate should be killed by this level of irradiation.
recBCD	Spot dilutions of λ gam$^-$ (UNIT 1.9) on a lawn of cells side by side with dilutions of λ gam$^+$. The *gam*$^-$ plaques should be almost as big as the *gam*$^+$ plaques.
hsdS$^-$	(1) Use the strain and a wild-type strain to plate out serial dilutions of a λ-like phage stock grown on an *hsdS*$^-$ or *hsdR*$^-$ host. If the phage stock came from an *hsdS*$^-$ host, then it should make plaques with 10^4 to 10^6 higher efficiency on the putative *hsdS*$^-$ host than on a wild-type host. If the plate stock came from an *hsdR*$^-$ (*hsdS*$^+$ *hsdM*$^+$) host, it should make plaques with the same efficiency on both strains. (2) Suspend one of the fresh plaques from the putative *hsdS*$^-$ host in 1 ml lambda dilution buffer. Titer this suspension on the putative *hsdS*$^-$ strain and on a wild-type strain. The suspension should make plaques at 10^4 to 10^6 higher efficiency on the *hsdS*$^-$ strain than on the wild-type strain. One plaque contains ~10^5 phage.
hsdR$^-$ (*hsdS*$^+$ *hsdM*$^+$)	(1) Perform step 1 described above, using a plate stock made from an *hsdS*$^-$ host. (2) Suspend one of the fresh plaques in 1 ml lambda dilution buffer. Titer this suspension on the putative *hsdR*$^-$ strain and on a wild-type strain. This suspension should make plaques with the same efficiency on the *hsdR*$^-$ strain as on a wild-type strain.
dam	Transform the strain and a wild-type strain with a plasmid that contains recognition sites for the enzymes *Mbo*I or *Bcl*I. Prepare plasmid DNA from both strains and verify that plasmid DNA isolated from the *dam*$^-$ strain is sensitive to digestion by the enzyme.
dcm	Transform the strain and a wild-type strain with a plasmid that contains recognition sites for *Scr*FI. Prepare plasmid DNA from both strains to verify that only plasmid DNA from the *dcm* strain is fully sensitive to digestion by the enzyme. Half of the *Scr*FI sites will be cut even when the DNA is *dcm*-methylated.
lon	Streak LB plate for single colonies. Also streak a control plate of a wild-type strain. Incubate at 37°C. Colonies of the *lon*$^-$ strain should be larger, glistening, and mucoidal.

[a]Commonly used protocols in this table are media preparation (UNIT 1.1), streaking and replicating a plate (UNIT 1.3), and growing lambda-derived vectors (UNIT 1.12).
[b]Encodes omega fragment of β-galactosidase.

Table 1.4.5 Commonly Used *Escherichia coli* Strains

Strain[a]	Genotype	Reference[b]
BNN102[c]	C600 *hflA150 chr::*Tn*10 mcrA1 mcrB*	Young and Davis, 1983*
BW313[d]	*Hfr lysA⁻ dut ung thi-1 recA spoT1*	Kunkel et al., 1987*†
C600	*thi-1 thr-1 leuB6 lacY1 tonA21 supE44 mcrA*	Appleyard, 1954*‡
CJ236[d]	*dut1 ung1 thi-1 relA1*/pCJ105 (Cm^r)	Kunkel et al., 1987*; Joyce and Grindley, 1984†
DH1	*recA1 endA1 thi-1 hsdR17 supE44 gyrA96* (Nal^r) *relA1*	Hanahan, 1983*; D. Hanahan, pers. comm.†‡
DH5αF′[e]	F′/endA1 *hsdR17*($r_K^-m_K^+$) *supE44 thi-1 recA1 gyrA* (Nal^r) *relA1* Δ*(lacZYA-argF)*$_{U169}$ (π*80lacZ*Δ*M15)*	See DH1 references
DK1	*hsdR2 hsdM⁺ hsdS⁺ araD139* Δ*(ara-leu)*$_{7697}$ Δ*(lac)*$_{X74}$ *galU galK* rpsL (Str^r) *mcrA mcrB1* Δ*(srl-recA)*$_{306}$	D. Kurnit and B. Seed, pers. comm.†‡
ER1451	F′ *traD36 proAB lacI*^{q} Δ*(lacZ)M15/endA gyrA96 thi-1 hsdR2* (or *hsdR17*) *supE44* Δ*(lac-proAB) mcrB1 mcrA*	Raleigh et al., 1988†‡
HB101[f]	*pro leu thi lacY hsdS*$_B$*20 endA recA rpsL20* (Str^r) *ara-14 galK2 xyl-5 mtl-1 supE44 mcrB*$_B$	Boyer and Roulland-Dussoix, 1969*†‡
JM101[g]	F′ *traD36 proA⁺ proB⁺ lacI*^{q} *lacZ*Δ*M15/supE thi* Δ*(lac-proAB)*	Yanisch-Perron et al., 1985*†‡
JM105[g]	F′ *traD36 proA⁺ proB⁺ lacI*^{q} *lacZ*Δ*M15/*Δ*(lac-pro)*$_{X111}$ *thi rpsL* (Str^r) *endA sbcB supE hsdR*	See JM101 references
JM107[g]	F′ *traD36 proA⁺ proB⁺ lacI*^{q} *lacZ*Δ*M15/endA1 gyrA96* (Nal^r) *thi hsdR17 supE44 relA1* Δ*(lac-proAB) mcrA*	See JM101 references
JM109[g]	F′ traD36 proA⁺ proB⁺ lacI^q *lacZ*Δ*M15/recA1 endA1 gyrA96* (Nal^r) *thi hsdR17 supE44 relA1* Δ*(lac-proAB) mcrA*	See JM101 references
KM392	*hsdR514*($r_K^-m_K^+$) *supE44 supF58 lacY galK2 galT22 metB1 trp55 mcrA* Δ*lac*$_{U169}$ *proC::*Tn*5*	T. St. John, pers. comm.†; K. Moore‡
LE392	*hsdR514*($r_K^-m_K^+$) *supE44 supF58 lacY galK2 galT22 metB1 trp55 mcrA*	Borck, et al., 1976*; N. Murray, pers. comm.†; L. Enquist‡
MC1061	*hsdR2 hsdM⁺ hsdS⁺ araD139* Δ*(ara-leu)*$_{7697}$ Δ*(lac)*$_{X74}$ *galE15 galK16 rpsL* (Str^r) *mcrA mcrB1*	Casadaban and Cohen, 1980*; M. Casadaban†‡

continued

the *E. coli* restriction mechanisms (Bickle, 1982). It attacks DNA that carries the site:

5′ A ^{m}A C N N N N N N G T G C 3′
3′ T T G N N N N N N C ^{m}A C G 5′

and results in double-strand cleavage at a variable distance from the site, leading eventually to degradation of the resulting fragments. DNA is not attacked if it lacks the site, or if the site is present but methylated at the adenines shown (^{m}A).

The *Eco*K enzyme is both genetically and enzymatically complex. The HsdR, HsdM, and HsdS subunits are required for restriction of an unmethylated substrate. The same complex will methylate the same substrate, but at a very slow rate, so that an unmethylated target rarely survives. A substrate methylated on only one strand (hemimethylated) will be methylated on the other strand by the three-protein complex, but will not be cut. HsdM and HsdS together can methylate either an unmethylated or a hemimethylated substrate. The three-protein complex is inactive for restriction if any of the three subunits is defective, but can still methylate if HsdR is defective.

In summary, a strain defective in the *hsdR* gene is described as having the phenotype HsdR⁻M⁺ (or, equivalently, *Eco*K R⁻M⁺ or $R_K^-M_K^+$; see box): it will methylate newly introduced DNA but will not restrict it. However, a strain defective in either *hsdM* or *hsdS* will neither restrict nor methylate, and has the phenotype HsdR⁻M⁻ (or *Eco*K R⁻M⁻ or $R_K^-M_K^-$).

In contrast with *Eco*K, the other three restriction systems of *E. coli* K-12—*mcrA*, *mcrB*, and *mrr*—specifically attack DNA that *is* methylated at particular sequences, rather

Table 1.4.5 Commonly Used *Escherichia coli* Strains, continued

Strain[c]	Genotype	Reference[b]
MM294	*endA thiA hsdR17 supE44*	Backman et al., 1976*; M. Meselson†‡
NM539[h]	*supF hsdR* (P2*cox*3)	Frischauf et al., 1983*; Lindahl and Sunshine, 1972†; N. Murray‡
P2392	*hsdR514*($r_K^- m_K^+$) *supE44 supF58 lacY galK2 galT22 metB1 trp55 mcrA* (P2)	L. Klickstein, pers. comm.†
Q359	*hsdR*⁻ *hsdM*⁺ *supE tonA* (φ80ʳ) (P2)	Karn et al., 1980*†‡
Y1088[i]	*supE supF metB trpR hsdR*⁻ *hsdM*⁺ *tonA21 strA* Δlac_{U169} *mcrA proC*::Tn*5* / pMC9	Huynh et al., 1985*; Miller et al., 1984†; R. Young‡; M. Calos (pMC9)‡
Y1089[i]	Δlac_{U169} *proA*⁺ Δ(*lon*) *araD139 strA hflA150 chr*::Tn*10* / pMC9	See Y1088 references
Y1090[i]	Δlac_{U169} *proA*⁺ Δ(*lon*) *araD139 strA supF trpC22*::Tn*10 mcrA* / pMC9	See Y1088 references

[a]The original *E. coli* K-12 strain was a λ lysogen, but most K-12 derivatives in common use have been cured of the prophage. By convention, all the other genes in these strains are presumed to be wild-type except for the genotype markers noted in the second column.

[b]Reference for all *mcr* and *mrr* genotypes is Raleigh et al., 1988. Specific information for each strain can be found as indicated by:
- * reference for genotype of strain;
- † source of additional genotype information;
- ‡ thought to be responsible for original strain construction.

[c]BNN102 is also known as C500 *hflA*.

[d]Both CJ236 and BW313 are commonly used in oligonucleotide-directed mutagenesis. pCJ105, the plasmid CJ236 carries, is not relevant for this application.

[e]Three strains are in circulation. DH5 is a derivative of DH1 that transforms at slightly higher efficiency. DH5α and DH5αF′ are derivatives that carry a deletion of the lac operon and a Φ80 prophage that directs synthesis of the omega fragment of β-galactosidase. DH5αF′ carries an F′ factor as well. DH5α and DH5αF′ are proprietary strains and the cells are prepared in some way that allows them to be transformed with slightly higher efficiency than DH5.

[f]In this strain, the area of the chromosome that contains the *hsd* genes was derived from the related B strain of *E. coli*.

[g]The continued presence of the F′ factor in JM strains can be insured by starting cultures only from single colonies grown on minimal plates that do not contain proline. These strains encode the omega fragment of *lacZ* and are frequently used with vectors that direct the synthesis of the *lacZ* alpha fragment. These strains are frequently used with M13 vectors for DNA sequencing (*UNITS 1.14, 1.15, & 7.4*).

[h]It is not known whether this strain has markers other than those listed.

[i]pMC9, the plasmid in the Y strains listed here, directs the synthesis of large amounts of *lac* repressor. It also confers resistance to tetracycline and ampicillin (Lebrowski et al., 1984, *EMBO. J.* 3:3117-3121).

DNA that is not. Either methylated cytosine residues or methylated adenine residues can create problems (see below).

The action of either *mcrA* or *mcrB* reduces the number of clones recovered from libraries made with genomic DNA from other organisms, and leads to bias against recovery of specific fragments from those libraries (Raleigh et al., 1988; Whittaker et al., 1988; Woodcock et al., 1988, 1989; *mrr* has not been tested). For McrB there is evidence that a nuclease is responsible for these effects (E. Sutherland and E.A. Raleigh, unpublished observation), but no such evidence is available for the other two systems.

Even without biochemical characterization, something can be said of the recognition sites for these systems. McrA restricts DNA modified by the *Hpa*II (5′ C ᵐCGG) methylase and possibly other methylases. McrB restricts DNA modified by any one of 14 other modification methylases, which led to the suggestion that the McrB recognition site is 5′ G ᵐC (Raleigh and Wilson, 1986). Mrr restricts DNA modified by the *Hha*II (5′ G ᵐA N T C) or *Pst*I (5′ C T G C ᵐA G) methylases, but *not* that modified by the *Eco*RI methylase, among others (Heitman and Model, 1987).

Many commonly used *E. coli* strains are McrA⁻, including (from Table 1.4.5) BNN102 (also known as C600*hflA*), C600, JM107, JM109, LE392, Y1088, and Y1090. Of the strains listed in Table 1.4.5, only BNN102, HB101, and MC1061 are McrB⁻; and only HB101 is Mrr⁻ (see also Raleigh et al., 1988).

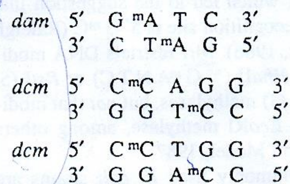
A strain should be used which lacks the appropriate methyl-specific restriction system(s) when cloning genomic DNA from an organism containing methylated bases. All mammals and higher plants, and many prokaryo-

tes, contain methylcytosine (Ehrlich and Wang, 1981), so McrA⁻B⁻ strains should be used for libraries of DNA from these organisms. Bacteria and lower eukaryotes may contain methyladenine, so Mrr sensitivity should be considered. However, the important experimental organisms *Drosophila melanogaster* and *Saccharomyces cerevisiae* contain no detectable methylated bases.

In addition, any time DNA is methylated in vitro during a manipulation, an appropriate restriction-deficient host should be used as DNA recipient. Methylases are used to generate novel restriction enzyme specificities or to protect cDNA from subsequent digestion (see UNITS 3.1 & 5.6). For example, the *Alu*I methylase (M.*Alu*I) is sometimes used to protect *Hin*dIII sites. McrB will restrict DNA modified by M.*Alu*I.

Once the DNA introduced into *E. coli* has been replicated, the foreign methylation pattern will be lost (and the *E. coli* methylation pattern will be acquired) unless the clone carries a methylase activity. Once successfully introduced, clones can be freely transferred among Mcr⁺ Mrr⁺ *E. coli* strains, since the methylation pattern will no longer be foreign. It is important that the clone be passed through an HsdM⁺ strain before trying to introduce it into an HsdR⁺ strain.

The normal methylation pattern of *E. coli* DNA is the product of three methylases. The *Eco*K methylase modifies the sequences indicated above. The *dam* and *dcm* gene products are also methylases (Marinus, 1987). The recognition sites for these are:



dam	5′ G ^{m}A T C 3′	
	3′ C T ^{m}A G 5′	
dcm	5′ C ^{m}C A G G 3′	
	3′ G G T ^{m}C C 5′	
dcm	5′ C ^{m}C T G G 3′	
	3′ G G A ^{m}C C 5′	

These modifications will render DNA resistant or partially resistant to some restriction endonucleases used for in vitro work (see Table 3.1.1), such as *Mbo*I and *Bcl*I (for Dam-modified DNA) or *Eco*RII (for Dcm-modified DNA). The Dam and Dcm methylases are *not* associated with any *E. coli* restriction function. Loss of Dam and/or Dcm methylation will not make the DNA sensitive to *Eco*K restriction, although loss of K modification will. However, Dam and Dcm modification confer sensitivity to Mrr and Mcr analogues in *Streptomyces* species (MacNeil, 1988).

RECOMBINATION AND ITS EFFECTS ON CLONED DNA INSERTS

During propagation in *E. coli*, DNA inserted into vectors is sometimes rearranged by the proteins involved in DNA recombination. Fortunately, although the genetics and enzymology of recombination in *E. coli* are still not well understood, there are mutant strains available that can provide solutions to two common cloning problems.

Problem 1. The DNA contains dispersed repeated sequences. Recombination occurs between these repeated sequences, causing loss of pieces of the DNA (see sketch).

For plasmid libraries, this problem can be solved by propagating the DNA in a *recA*⁻ host, where homologous recombination does not occur. For libraries made using λ-derived vectors, the vector must also be recombination-defective (*red*). However, only about 30% to 50% of the cells are viable in such a strain, and libraries, particularly phage libraries, may be hard to propagate. Phage λ vectors that are *red* will not make high-titer lysates in *recA* strains, and *red gam* phage will not grow at all, unless the *recBCD* enzyme is also inactivated (see below). Many λ vectors are *red gam* to make use of the Spi⁻ selection or to make room for larger insert pieces (see UNIT 1.10).

Problem 2. The inserted DNA contains closely spaced inverted repeat sequences (palindromes or interrupted palindromes). Such stretches of DNA are not stably propagated in either phage or plasmid vectors. Available knowledge is consistent with the idea that large (>300-bp) palindromes can sometimes form an alternative, hairpin structure that resembles an intermediate found in normal recombination called a Holliday junction, and are then acted upon by the host recombination system in such a way that the hairpin is eliminated or made smaller.

There are strains of bacteria from which phage and plasmid clones containing palindromes are recovered at higher frequency. These bacteria have inactivated exonuclease V (ExoV; encoded by the *recB*, *recC*, and *recD* genes) or the SbcC product (encoded by the *sbcC* gene). Many strains permissive for palindromes have defects in *recB recC* combined with a defect in *sbcA* (which prob-