

Chapter 6

Cloning Vectors for *E. coli*

Chapter contents

CHAPTER CONTENTS

- 6.1 Cloning vectors based on *E. coli* plasmids
- 6.2 Cloning vectors based on M13 bacteriophage
- 6.3 Cloning vectors based on λ bacteriophage
- 6.4 λ and other high-capacity vectors enable genomic libraries to be constructed
- 6.5 Vectors for other bacteria

The basic experimental techniques involved in gene cloning have now been described. In Chapters 3, 4, and 5 we have seen how DNA is purified from cell extracts, how recombinant DNA molecules are constructed in the test tube, how DNA molecules are reintroduced into living cells, and how recombinant clones are distinguished. Now we must look more closely at the cloning vector itself, in order to consider the range of vectors available to the molecular biologist, and to understand the properties and uses of each individual type.

The greatest variety of cloning vectors exists for use with *E. coli* as the host organism. This is not surprising in view of the central role that this bacterium has played in basic research over the past 50 years. The tremendous wealth of information that exists concerning the microbiology, biochemistry, and genetics of *E. coli* has meant that virtually all fundamental studies of gene structure and function were initially carried out with this bacterium as the experimental organism. Even when a eukaryote is being studied, *E. coli* is still used as the workhorse for preparation of cloned DNA for sequencing, and for construction of recombinant genes that will subsequently be placed back in the eukaryotic host in order to study their function and expression. In recent years, gene cloning and molecular biological research have become mutually synergistic—breakthroughs in gene cloning have acted as a stimulus to research, and the needs of research have spurred on the development of new, more sophisticated cloning vectors.

In this chapter the most important types of *E. coli* cloning vector will be described, and their specific uses outlined. In Chapter 7, cloning vectors for yeast, fungi, plants, and animals will be considered.

6.1 Cloning vectors based on *E. coli* plasmids

The simplest cloning vectors, and the ones most widely used in gene cloning, are those based on small bacterial plasmids. A large number of different plasmid vectors are available for use with *E. coli*, many obtainable from commercial suppliers. They combine ease of purification with desirable properties such as high transformation efficiency, convenient selectable markers for transformants and recombinants, and the ability to clone reasonably large (up to about 8 kb) pieces of DNA. Most “routine” gene cloning experiments make use of one or other of these plasmid vectors.

One of the first vectors to be developed was pBR322, which was introduced in Chapter 5 to illustrate the general principles of transformant selection and recombinant identification (p. 77). Although pBR322 lacks the more sophisticated features of the newest cloning vectors, and so is no longer used extensively in research, it still illustrates the important, fundamental properties of any plasmid cloning vector. We will therefore begin our study of *E. coli* vectors by looking more closely at pBR322.

6.1.1 The nomenclature of plasmid cloning vectors

The name “pBR322” conforms with the standard rules for vector nomenclature:

- “p” indicates that this is indeed a plasmid.
- “BR” identifies the laboratory in which the vector was originally constructed (BR stands for Bolivar and Rodriguez, the two researchers who developed pBR322).
- “322” distinguishes this plasmid from others developed in the same laboratory (there are also plasmids called pBR325, pBR327, pBR328, etc.).

6.1.2 The useful properties of pBR322

The genetic and physical map of pBR322 (Figure 6.1) gives an indication of why this plasmid was such a popular cloning vector.

The first useful feature of pBR322 is its size. In Chapter 2 it was stated that a cloning vector ought to be less than 10 kb in size, to avoid problems such as DNA breakdown

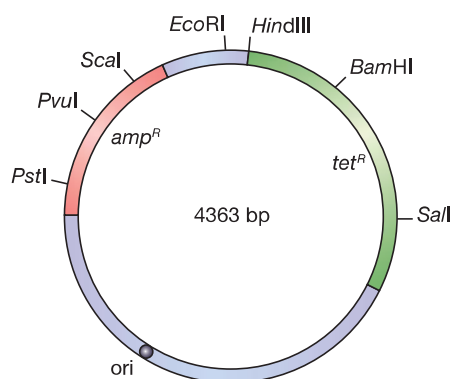


Figure 6.1

A map of pBR322 showing the positions of the ampicillin resistance (amp^R) and tetracycline resistance (tet^R) genes, the origin of replication (ori) and some of the most important restriction sites.

during purification. pBR322 is 4363 bp, which means that not only can the vector itself be purified with ease, but so can recombinant DNA molecules constructed with it. Even with 6 kb of additional DNA, a recombinant pBR322 molecule is still a manageable size.

The second feature of pBR322 is that, as described in Chapter 5, it carries two sets of antibiotic resistance genes. Either ampicillin or tetracycline resistance can be used as a selectable marker for cells containing the plasmid, and each marker gene includes unique restriction sites that can be used in cloning experiments. Insertion of new DNA into pBR322 that has been restricted with *Pst*I, *Pvu*I, or *Sca*I inactivates the *amp*^R gene, and insertion using any one of eight restriction endonucleases (notably *Bam*HI and *Hind*III) inactivates tetracycline resistance. This great variety of restriction sites that can be used for insertional inactivation means that pBR322 can be used to clone DNA fragments with any of several kinds of sticky end.

A third advantage of pBR322 is that it has a reasonably high copy number. Generally there are about 15 molecules present in a transformed *E. coli* cell, but this number can be increased, up to 1000–3000, by plasmid amplification in the presence of a protein synthesis inhibitor such as chloramphenicol (p. 39). An *E. coli* culture therefore provides a good yield of recombinant pBR322 molecules.

6.1.3 The pedigree of pBR322

The remarkable convenience of pBR322 as a cloning vector did not arise by chance. The plasmid was in fact designed in such a way that the final construct would possess these desirable properties. An outline of the scheme used to construct pBR322 is shown in Figure 6.2a. It can be seen that its production was a tortuous business that required full and skilful use of the DNA manipulative techniques described in Chapter 4. A summary of the result of these manipulations is provided in Figure 6.2b, from which it can be seen that pBR322 comprises DNA derived from three different naturally occurring plasmids. The *amp*^R gene originally resided on the plasmid R1, a typical antibiotic resistance plasmid that occurs in natural populations of *E. coli* (p. 17). The *tet*^R gene is derived from R6-5, a second antibiotic-resistant plasmid. The replication origin of pBR322, which directs multiplication of the vector in host cells, is originally from pMB1, which is closely related to the colicin-producing plasmid ColE1 (p. 17).

6.1.4 More sophisticated *E. coli* plasmid cloning vectors

pBR322 was developed in the late 1970s, the first research paper describing its use being published in 1977. Since then many other plasmid cloning vectors have been constructed, the majority of these derived from pBR322 by manipulations similar to those summarized in Figure 6.2a. One of the first of these was pBR327, which was produced by removing a 1089 bp segment from pBR322. This deletion left the *amp*^R and *tet*^R genes intact, but changed the replicative and conjugative abilities of the resulting plasmid. As a result, pBR327 differs from pBR322 in two important ways:

- pBR327 has a higher copy number than pBR322, being present at about 30–45 molecules per *E. coli* cell. This is not of great relevance as far as plasmid yield is concerned, as both plasmids can be amplified to copy numbers greater than 1000. However, the higher copy number of pBR327 in normal cells makes this vector more suitable if the aim of the experiment is to study the function of the cloned gene. In these cases gene dosage becomes important, because the more copies there

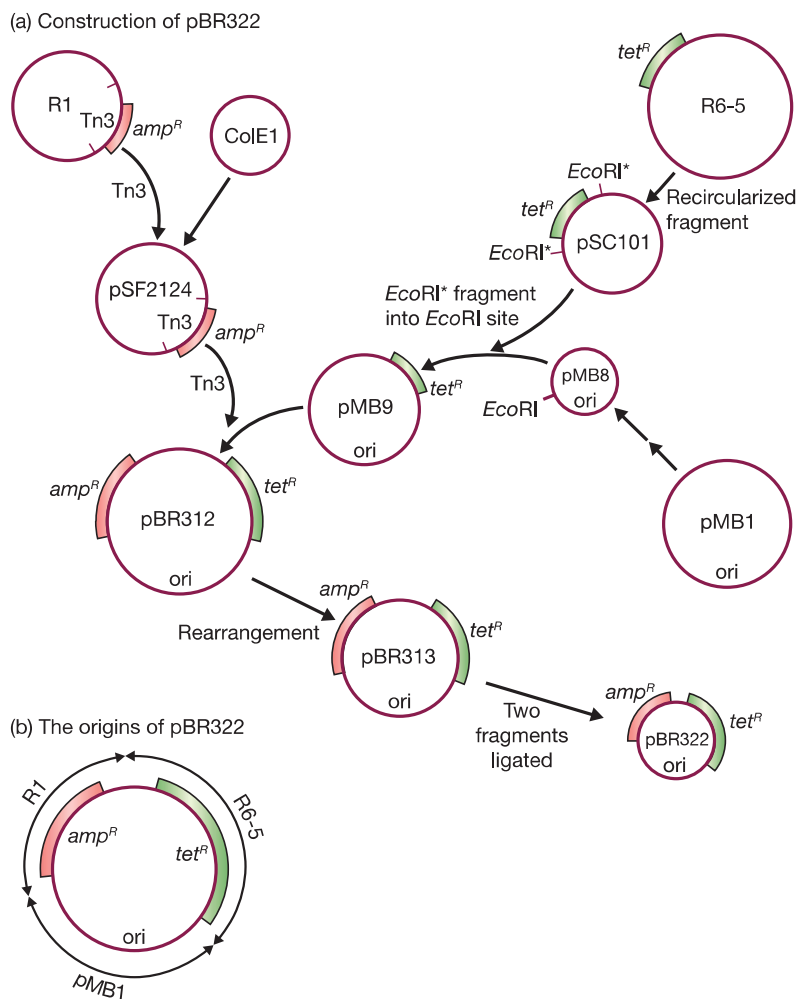


Figure 6.2

The pedigree of pBR322. (a) The manipulations involved in construction of pBR322. (b) A summary of the origins of pBR322.

are of a cloned gene, the more likely it is that the effect of the cloned gene on the host cell will be detectable. pBR327, with its high copy number, is therefore a better choice than pBR322 for this kind of work.

- The deletion also destroys the conjugative ability of pBR322, making pBR327 a non-conjugative plasmid that cannot direct its own transfer to other *E. coli* cells. This is important for **biological containment**, averting the possibility of a recombinant pBR327 molecule escaping from the test tube and colonizing bacteria in the gut of a careless molecular biologist. In contrast, pBR322 could theoretically be passed to natural populations of *E. coli* by conjugation, though in fact pBR322 also has safeguards (though less sophisticated ones) to minimize the chances of this happening. pBR327 is, however, preferable if the cloned gene is potentially harmful should an accident occur.

Although pBR327, like pBR322, is no longer widely used, its properties have been inherited by most of today's modern plasmid vectors. There are a great number of these, and it would be pointless to attempt to describe them all. Two additional examples will suffice to illustrate the most important features.

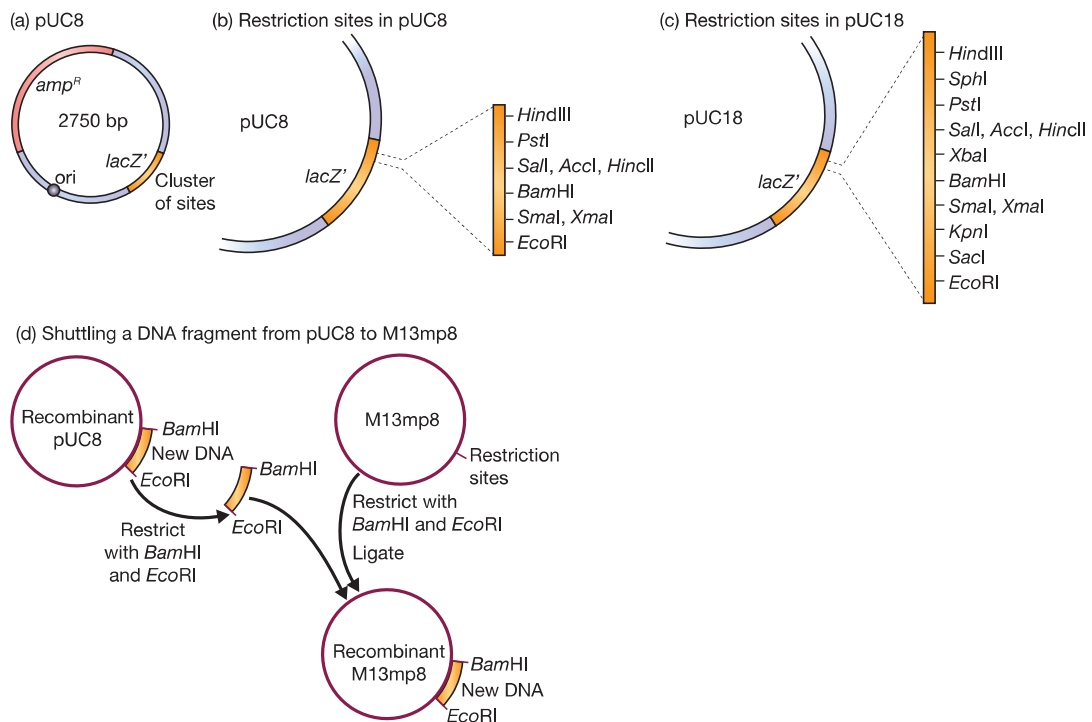


Figure 6.3

The pUC plasmids. (a) The structure of pUC8. (b) The restriction site cluster in the *lacZ'* gene of pUC8. (c) The restriction site cluster in pUC18. (d) Shuttling a DNA fragment from pUC8 to M13mp8.

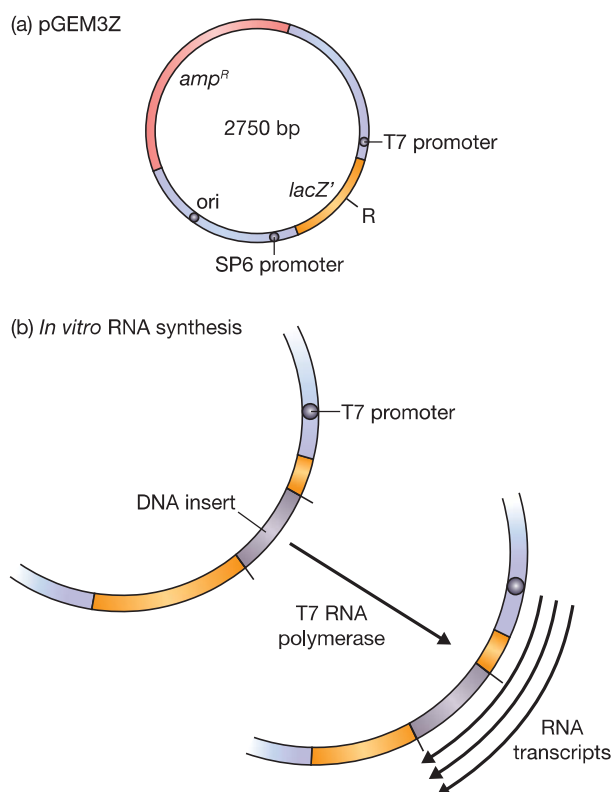
pUC8—a Lac selection plasmid

This vector was mentioned in Chapter 5 when identification of recombinants by insertional inactivation of the β -galactosidase gene was described (p. 79). pUC8 (Figure 6.3a) is descended from pBR322, although only the replication origin and the *amp^R* gene remain. The nucleotide sequence of the *amp^R* gene has been changed so that it no longer contains the unique restriction sites: all these cloning sites are now clustered into a short segment of the *lacZ'* gene carried by pUC8.

pUC8 has three important advantages that have led to it becoming one of the most popular *E. coli* cloning vectors. The first of these is fortuitous: the manipulations involved in construction of pUC8 were accompanied by a chance mutation, within the origin of replication, which results in the plasmid having a copy number of 500–700 even before amplification. This has a significant effect on the yield of cloned DNA obtainable from *E. coli* cells transformed with recombinant pUC8 plasmids.

The second advantage is that identification of recombinant cells can be achieved by a single step process, by plating onto agar medium containing ampicillin plus X-gal (p. 79). With both pBR322 and pBR327, selection of recombinants is a two-step procedure, requiring replica plating from one antibiotic medium to another (p. 78). A cloning experiment with pUC8 can therefore be carried out in half the time needed with pBR322 or pBR327.

The third advantage of pUC8 lies with the clustering of the restriction sites, which allows a DNA fragment with two different sticky ends (say *EcoRI* at one end and *BamHI* at the other) to be cloned without resorting to additional manipulations such as linker attachment (Figure 6.3b). Other pUC vectors carry different combinations of restriction

**Figure 6.4**

pGEM3Z. (a) Map of the vector. (b) *In vitro* RNA synthesis. R = cluster of restriction sites for *EcoRI*, *SacI*, *KpnI*, *AvaI*, *SmaI*, *BamHI*, *XbaI*, *SalI*, *AccI*, *HincII*, *PstI*, *SphI*, and *HindIII*.

sites and provide even greater flexibility in the types of DNA fragment that can be cloned (Figure 6.3c). Furthermore, the restriction site clusters in these vectors are the same as the clusters in the equivalent M13mp series of vectors (p. 95). DNA cloned into a member of the pUC series can therefore be transferred directly to its M13mp counterpart, enabling the cloned gene to be obtained as single-stranded DNA (Figure 6.3d).

pGEM3Z—*in vitro* transcription of cloned DNA

pGEM3Z (Figure 6.4a) is very similar to a pUC vector: it carries the *amp^R* and *lacZ'* genes, the latter containing a cluster of restriction sites, and it is almost exactly the same size. The distinction is that pGEM3Z has two additional short pieces of DNA, each of which acts as the recognition site for attachment of an RNA polymerase enzyme. These two **promoter** sequences lie on either side of the cluster of restriction sites used for introduction of new DNA into the pGEM3Z molecule. This means that if a recombinant pGEM3Z molecule is mixed with purified RNA polymerase in the test tube, transcription occurs and RNA copies of the cloned fragment are synthesized (Figure 6.4b). The RNA that is produced could be used as a hybridization probe (p. 133), or might be required for experiments aimed at studying RNA processing (e.g., the removal of introns) or protein synthesis.

The promoters carried by pGEM3Z and other vectors of this type are not the standard sequences recognized by the *E. coli* RNA polymerase. Instead, one of the promoters is specific for the RNA polymerase coded by T7 bacteriophage and the other for the RNA polymerase of SP6 phage. These RNA polymerases are synthesized during infection of *E. coli* with one or other of the phages and are responsible for transcribing the phage genes. They are chosen for use in *in vitro* transcription as they are very active enzymes – remember that the entire lytic infection cycle takes only 20 minutes (p. 18), so the

phage genes must be transcribed very quickly. These polymerases are able to synthesize 1–2 mg of RNA per minute, substantially more than can be produced by the standard *E. coli* enzyme.

6.2 Cloning vectors based on M13 bacteriophage

The most essential requirement for any cloning vector is that it has a means of replicating in the host cell. For plasmid vectors this requirement is easy to satisfy, as relatively short DNA sequences are able to act as plasmid origins of replication, and most, if not all, of the enzymes needed for replication are provided by the host cell. Elaborate manipulations, such as those that resulted in pBR322 (see Figure 6.2a), are therefore possible so long as the final construction has an intact, functional replication origin.

With bacteriophages such as M13 and λ , the situation as regards replication is more complex. Phage DNA molecules generally carry several genes that are essential for replication, including genes coding for components of the phage protein coat and phage-specific DNA replicative enzymes. Alteration or deletion of any of these genes will impair or destroy the replicative ability of the resulting molecule. There is therefore much less freedom to modify phage DNA molecules, and generally phage cloning vectors are only slightly different from the parent molecule.

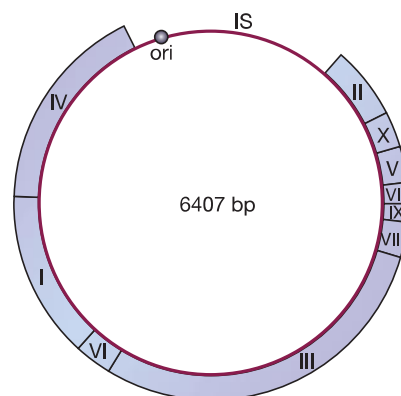
6.2.1 How to construct a phage cloning vector

The problems in constructing a phage cloning vector are illustrated by considering M13. The normal M13 genome is 6.4 kb in length, but most of this is taken up by ten closely packed genes (Figure 6.5), each essential for the replication of the phage. There is only a single 507-nucleotide intergenic sequence into which new DNA could be inserted without disrupting one of these genes, and this region includes the replication origin which must itself remain intact. Clearly there is only limited scope for modifying the M13 genome.

The first step in construction of an M13 cloning vector was to introduce the *lacZ'* gene into the intergenic sequence. This gave rise to M13mp1, which forms blue plaques on X-gal agar (Figure 6.6a). M13mp1 does not possess any unique restriction sites in the *lacZ'* gene. It does, however, contain the hexanucleotide GGATTC near the start of the gene. A single nucleotide change would make this GAATTC, which is an *EcoRI* site. This alteration was carried out using *in vitro* mutagenesis (p. 200), resulting in M13mp2

Figure 6.5

The M13 genome, showing the positions of genes I to X.



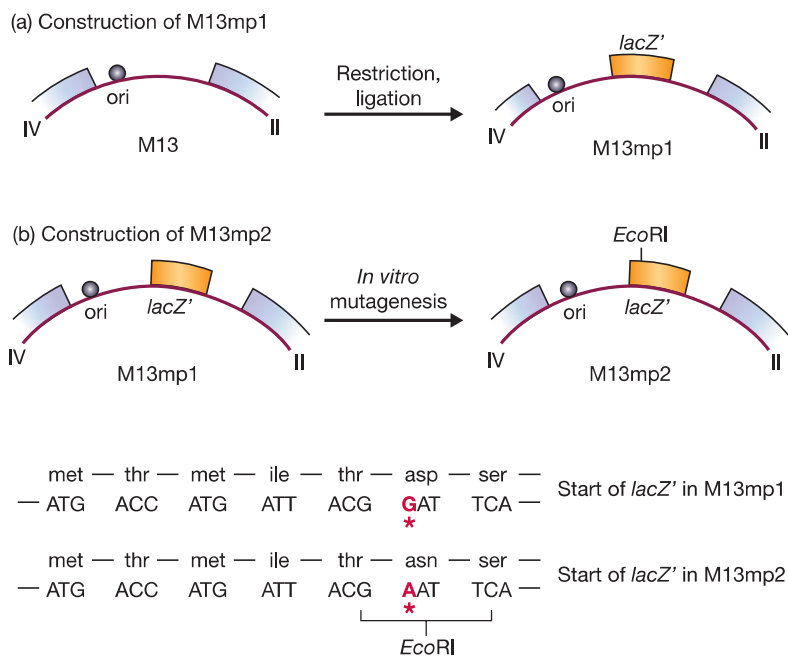


Figure 6.6

Construction of (a) M13mp1, and (b) M13mp2 from the wild-type M13 genome.

(Figure 6.6b). M13mp2 has a slightly altered *lacZ'* gene (the sixth codon now specifies asparagine instead of aspartic acid), but the β-galactosidase enzyme produced by cells infected with M13mp2 is still perfectly functional.

The next step in the development of M13 vectors was to introduce additional restriction sites into the *lacZ'* gene. This was achieved by synthesizing in the test tube a short oligonucleotide, called a **polylinker**, which consists of a series of restriction sites and has *EcoRI* sticky ends (Figure 6.7a). This polylinker was inserted into the *EcoRI* site of M13mp2, to give M13mp7 (Figure 6.7b), a more complex vector with four possible cloning sites (*EcoRI*, *Bam*HI, *Sal*I, and *Pst*I). The polylinker is designed so that it does

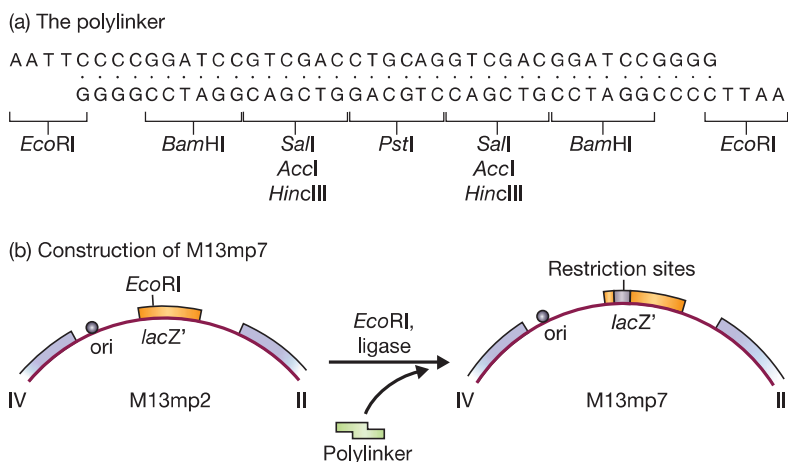


Figure 6.7

Construction of M13mp7: (a) the polylinker, and (b) its insertion into the *EcoRI* site of M13mp2. Note that the *SalI* restriction sites are also recognized by *AccI* and *HinII*.

not totally disrupt the *lacZ'* gene: a reading frame is maintained throughout the polylinker, and a functional, though altered, β -galactosidase enzyme is still produced.

The most sophisticated M13 vectors have more complex polylinkers inserted into the *lacZ'* gene. An example is M13mp8, which has the same series of restriction sites as the plasmid pUC8 (p. 92). As with the plasmid vector, one advantage of M13mp8 is its ability to take DNA fragments with two different sticky ends.

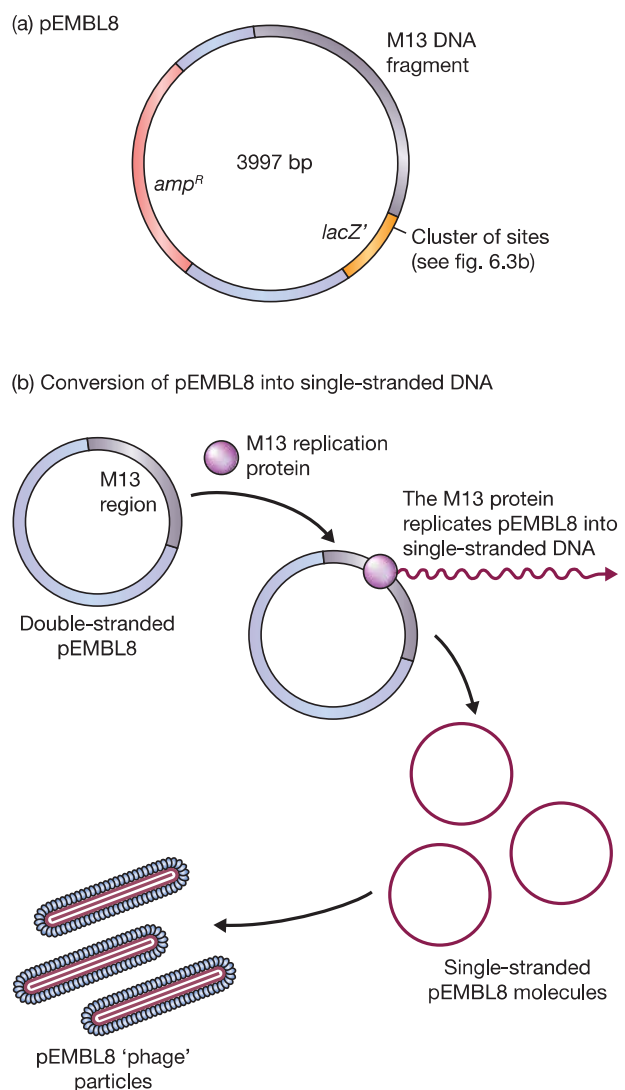
6.2.2 Hybrid plasmid–M13 vectors

Although M13 vectors are very useful for the production of single-stranded versions of cloned genes, they do suffer from one disadvantage. There is a limit to the size of DNA fragment that can be cloned with an M13 vector, with 1500 bp generally being looked on as the maximum capacity, though fragments up to 3 kb have occasionally been cloned. To get around this problem a number of hybrid vectors (“**phagemids**”) have been developed by combining a part of the M13 genome with plasmid DNA.

An example is provided by pEMBL8 (Figure 6.8a), which was made by transferring into pUC8 a 1300 bp fragment of the M13 genome. This piece of M13 DNA contains

Figure 6.8

pEMBL8: a hybrid plasmid–M13 vector that can be converted into single-stranded DNA.



the signal sequence recognized by the enzymes that convert the normal double-stranded M13 molecule into single-stranded DNA before secretion of new phage particles. This signal sequence is still functional even though detached from the rest of the M13 genome, so pEMBL8 molecules are also converted into single-stranded DNA and secreted as defective phage particles (Figure 6.8b). All that is necessary is that the *E. coli* cells used as hosts for a pEMBL8 cloning experiment are subsequently infected with normal M13 to act as a **helper phage**, providing the necessary replicative enzymes and phage coat proteins. pEMBL8, being derived from pUC8, has the polylinker cloning sites within the *lacZ'* gene, so recombinant plaques can be identified in the standard way on agar containing X-gal. With pEMBL8, single-stranded versions of cloned DNA fragments up to 10 kb in length can be obtained, greatly extending the range of the M13 cloning system.

6.3 Cloning vectors based on λ bacteriophage

Two problems had to be solved before λ -based cloning vectors could be developed:

- The λ DNA molecule can be increased in size by only about 5%, representing the addition of only 3 kb of new DNA. If the total size of the molecule is more than 52 kb, then it cannot be packaged into the λ head structure and infective phage particles are not formed. This severely limits the size of a DNA fragment that can be inserted into an unmodified λ vector (Figure 6.9a).
- The λ genome is so large that it has more than one recognition sequence for virtually every restriction endonuclease. Restriction cannot be used to cleave the normal λ molecule in a way that will allow insertion of new DNA, because the molecule would be cut into several small fragments that would be very unlikely to re-form a viable λ genome on religation (Figure 6.9b).

In view of these difficulties it is perhaps surprising that a wide variety of λ cloning vectors have been developed, their primary use being to clone large pieces of DNA, from 5 to 25 kb, much too big to be handled by plasmid or M13 vectors.

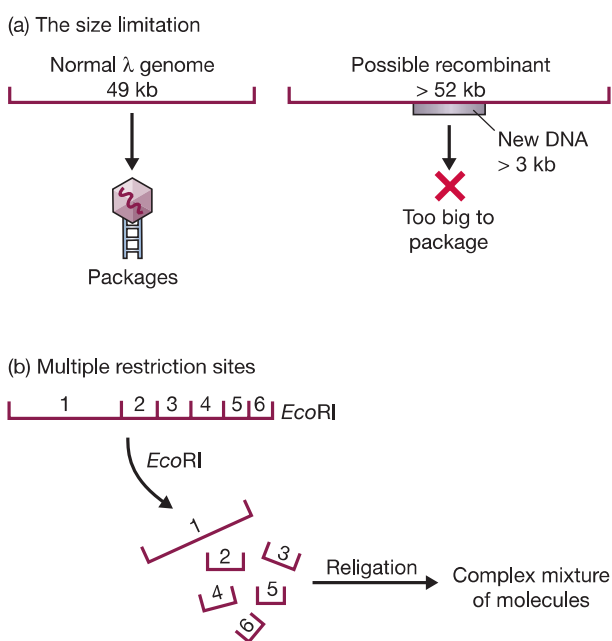
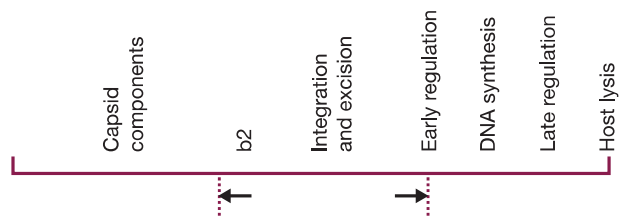


Figure 6.9

The two problems that had to be solved before λ cloning vectors could be developed. (a) The size limitation placed on the λ genome by the need to package it into the phage head. (b) λ DNA has multiple recognition sites for almost all restriction endonucleases.

Figure 6.10

The λ genetic map, showing the position of the main non-essential region that can be deleted without affecting the ability of the phage to follow the lytic infection cycle. There are other, much shorter non-essential regions in other parts of the genome.



6.3.1 Segments of the λ genome can be deleted without impairing viability

The way forward for the development of λ cloning vectors was provided by the discovery that a large segment in the central region of the λ DNA molecule can be removed without affecting the ability of the phage to infect *E. coli* cells. Removal of all or part of this non-essential region, between positions 20 and 35 on the map shown in Figure 2.9, decreases the size of the resulting λ molecule by up to 15 kb. This means that as much as 18 kb of new DNA can now be added before the cut-off point for packaging is reached (Figure 6.10).

This “non-essential” region in fact contains most of the genes involved in integration and excision of the λ prophage from the *E. coli* chromosome. A deleted λ genome is therefore non-lysogenic and can follow only the lytic infection cycle. This in itself is desirable for a cloning vector as it means induction is not needed before plaques are formed (p. 40).

6.3.2 Natural selection can be used to isolate modified λ that lack certain restriction sites

Even a deleted λ genome, with the non-essential region removed, has multiple recognition sites for most restriction endonucleases. This is a problem that is often encountered when a new vector is being developed. If just one or two sites need to be removed, then the technique of *in vitro* mutagenesis (p. 200) can be used. For example, an *EcoRI* site, GAATTC, could be changed to GGATTC, which is not recognized by the enzyme. However, *in vitro* mutagenesis was in its infancy when the first λ vectors were under development, and even today would not be an efficient means of changing more than a few sites in a single molecule.

Instead, natural selection was used to provide strains of λ that lack the unwanted restriction sites. Natural selection can be brought into play by using as a host an *E. coli* strain that produces *EcoRI*. Most λ DNA molecules that invade the cell are destroyed by this restriction endonuclease, but a few survive and produce plaques. These are mutant phages, from which one or more *EcoRI* sites have been lost spontaneously (Figure 6.11). Several cycles of infection will eventually result in λ molecules that lack all or most of the *EcoRI* sites.

6.3.3 Insertion and replacement vectors

Once the problems posed by packaging constraints and by the multiple restriction sites had been solved, the way was open for the development of different types of λ -based cloning vectors. The first two classes of vector to be produced were λ **insertion** and λ **replacement** (or substitution) vectors.

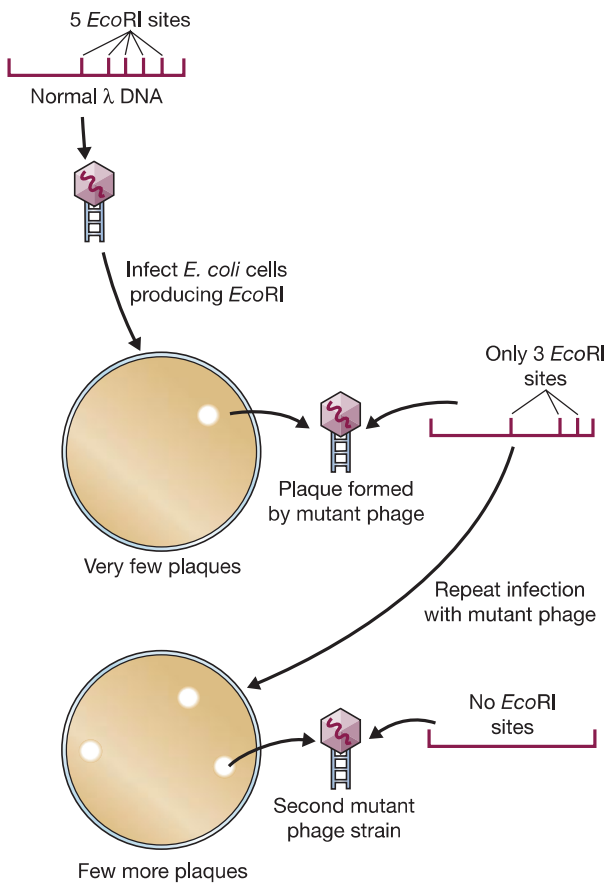


Figure 6.11

Using natural selection to isolate λ phage lacking *EcoRI* restriction sites.

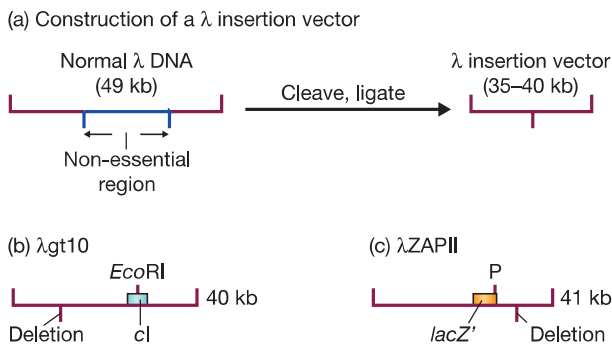


Figure 6.12

λ insertion vectors. P = polylinker in the *lacZ'* gene of λZAP II, containing unique restriction sites for *SacI*, *NotI*, *XbaI*, *SpeI*, *EcoRI*, and *XhoI*.

Insertion vectors

With an insertion vector (Figure 6.12a), a large segment of the non-essential region has been deleted, and the two arms ligated together. An insertion vector possesses at least one unique restriction site into which new DNA can be inserted. The size of the DNA fragment that an individual vector can carry depends, of course, on the extent to which the non-essential region has been deleted. Two popular insertion vectors are:

- **λgt10** (Figure 6.12b), which can carry up to 8 kb of new DNA, inserted into a unique *EcoRI* site located in the *cI* gene. Insertional inactivation of this gene means that recombinants are distinguished as clear rather than turbid plaques (p. 83).

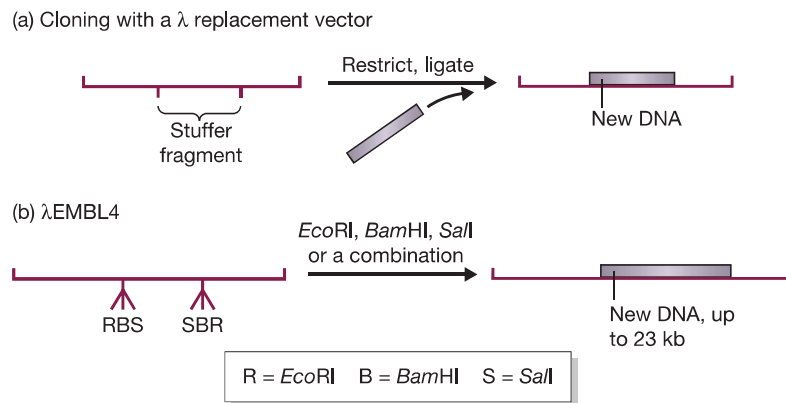


Figure 6.13

λ replacement vectors. (a) Cloning with a λ replacement vector. (b) Cloning with λ EMBL4.

- **λ ZAPII** (Figure 6.12c), with which insertion of up to 10 kb DNA into any of 6 restriction sites within a polylinker inactivates the *lacZ'* gene carried by the vector. Recombinants give clear rather than blue plaques on X-gal agar.

Replacement vectors

A λ replacement vector has two recognition sites for the restriction endonuclease used for cloning. These sites flank a segment of DNA that is replaced by the DNA to be cloned (Figure 6.13a). Often the replaceable fragment (or “**stuffer fragment**” in cloning jargon) carries additional restriction sites that can be used to cut it up into small pieces, so that its own re-insertion during a cloning experiment is very unlikely. Replacement vectors are generally designed to carry larger pieces of DNA than insertion vectors can handle. Recombinant selection is often on the basis of size, with non-recombinant vectors being too small to be packaged into λ phage heads (p. 84).

An example of a replacement vectors is:

- **λ EMBL4** (Figure 6.13b) can carry up to 20 kb of inserted DNA by replacing a segment flanked by pairs of *EcoRI*, *BamHI*, and *SalI* sites. Any of these three restriction endonucleases can be used to remove the stuffer fragment, so DNA fragments with a variety of sticky ends can be cloned. Recombinant selection with λ EMBL4 can be on the basis of size, or can utilize the Spi phenotype (p. 83).

6.3.4 Cloning experiments with λ insertion or replacement vectors

A cloning experiment with a λ vector can proceed along the same lines as with a plasmid vector—the λ molecules are restricted, new DNA is added, the mixture is ligated, and the resulting molecules used to transfect a competent *E. coli* host (Figure 6.14a). This type of experiment requires that the vector be in its circular form, with the *cos* sites hydrogen bonded to each other.

Although satisfactory for many purposes, a procedure based on transfection is not particularly efficient. A greater number of recombinants will be obtained if one or two refinements are introduced. The first is to use the linear form of the vector. When the linear form of the vector is digested with the relevant restriction endonuclease, the left and right arms are released as separate fragments. A recombinant molecule can

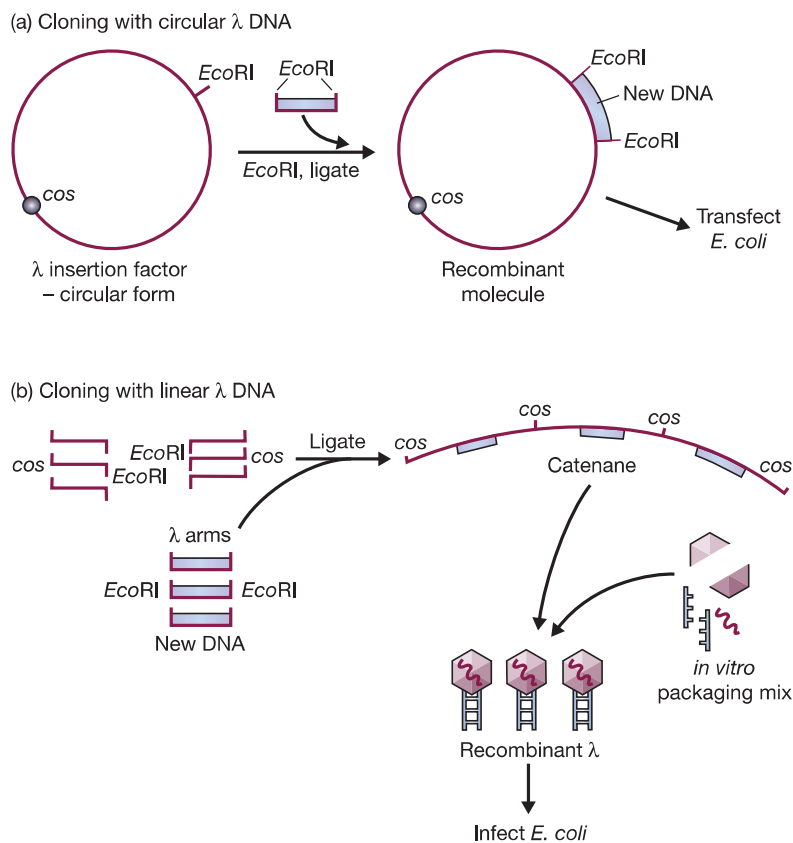


Figure 6.14

Different strategies for cloning with a λ vector. (a) Using the circular form of λ as a plasmid. (b) Using left and right arms of the λ genome, plus *in vitro* packaging, to achieve a greater number of recombinant plaques.

be constructed by mixing together the DNA to be cloned with the vector arms (Figure 6.14b). Ligation results in several molecular arrangements, including catenanes comprising left arm–DNA–right arm repeated many times (Figure 6.14b). If the inserted DNA is the correct size, then the *cos* sites that separate these structures will be the right distance apart for *in vitro* packaging (p. 81). Recombinant phage are therefore produced in the test tube and can be used to infect an *E. coli* culture. This strategy, in particular the use of *in vitro* packaging, results in a large number of recombinant plaques.

6.3.5 Long DNA fragments can be cloned using a cosmid

The final and most sophisticated type of λ -based vector is the **cosmid**. Cosmids are hybrids between a phage DNA molecule and a bacterial plasmid, and their design centers on the fact that the enzymes that package the λ DNA molecule into the phage protein coat need only the *cos* sites in order to function (p. 21). The *in vitro* packaging reaction works not only with λ genomes, but also with any molecule that carries *cos* sites separated by 37–52 kb of DNA.

A cosmid is basically a plasmid that carries a *cos* site (Figure 6.15a). It also needs a selectable marker, such as the ampicillin resistance gene, and a plasmid origin of replication, as cosmids lack all the λ genes and so do not produce plaques. Instead colonies are formed on selective media, just as with a plasmid vector.

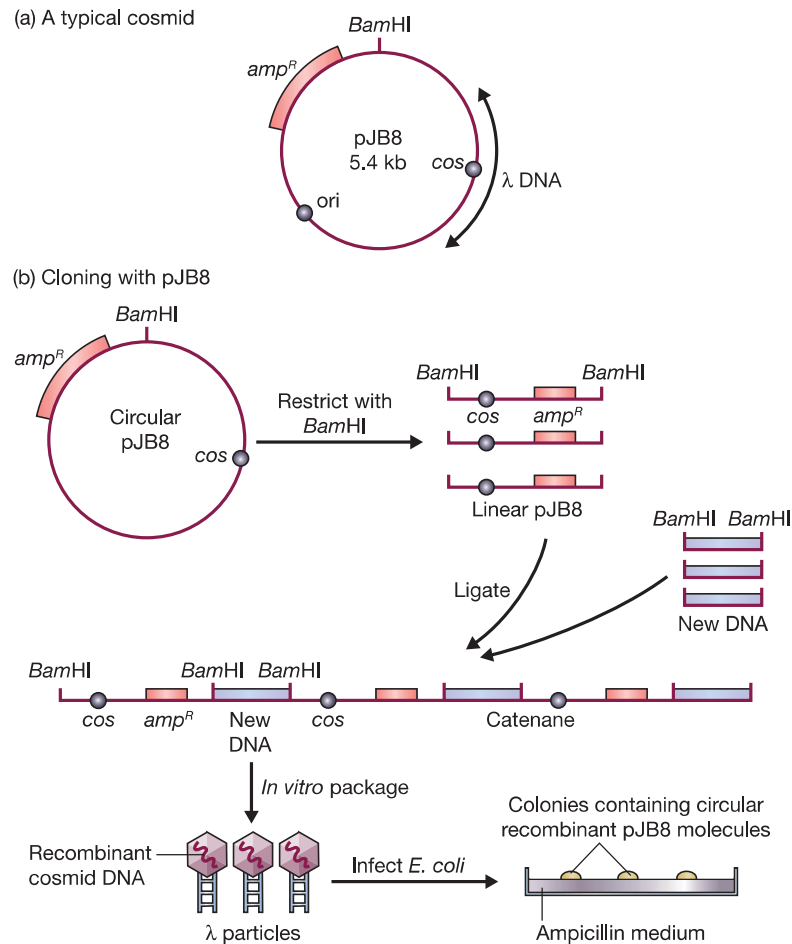


Figure 6.15

A typical cosmid and the way it is used to clone long fragments of DNA.

A cloning experiment with a cosmid is carried out as follows (Figure 6.15b). The cosmid is opened at its unique restriction site and new DNA fragments inserted. These fragments are usually produced by partial digestion with a restriction endonuclease, as total digestion almost invariably results in fragments that are too small to be cloned with a cosmid. Ligation is carried out so that catenanes are formed. Providing the inserted DNA is the right size, *in vitro* packaging cleaves the *cos* sites and places the recombinant cosmids in mature phage particles. These λ phage are then used to infect an *E. coli* culture, though of course plaques are not formed. Instead, infected cells are plated onto a selective medium and antibiotic-resistant colonies are grown. All colonies are recombinants, as non-recombinant linear cosmids are too small to be packaged into λ heads.

6.4 λ and other high-capacity vectors enable genomic libraries to be constructed

The main use of all λ -based vectors is to clone DNA fragments that are too long to be handled by plasmid or M13 vectors. A replacement vector, such as λ EMBL4, can carry up to 20 kb of new DNA, and some cosmids can manage fragments up to 40 kb. This

Table 6.1

Number of clones needed for genomic libraries of a variety of organisms.

SPECIES	GENOME SIZE (bp)	NUMBER OF CLONES*	
		17 kb FRAGMENTS [†]	35 kb FRAGMENTS [‡]
<i>E. coli</i>	4.6×10^6	820	410
<i>Saccharomyces cerevisiae</i>	1.8×10^7	3225	1500
<i>Drosophila melanogaster</i>	1.2×10^8	21,500	10,000
Rice	5.7×10^8	100,000	49,000
Human	3.2×10^9	564,000	274,000
Frog	2.3×10^{10}	4,053,000	1,969,000

*Calculated for a probability (p) of 95% that any particular gene will be present in the library.[†]Fragments suitable for a replacement vector such as λ EMBL4.[‡]Fragments suitable for a cosmid.

compares with a maximum insert size of about 8 kb for most plasmids and less than 3 kb for M13 vectors.

The ability to clone such long DNA fragments means that **genomic libraries** can be generated. A genomic library is a set of recombinant clones that contains all of the DNA present in an individual organism. An *E. coli* genomic library, for example, contains all the *E. coli* genes, so any desired gene can be withdrawn from the library and studied. Genomic libraries can be retained for many years, and propagated so that copies can be sent from one research group to another.

The big question is how many clones are needed for a genomic library? The answer can be calculated with the formula:

$$N = \frac{\ln(1 - p)}{\ln\left(1 - \frac{a}{b}\right)}$$

where N is the number of clones that are required, p is probability that any given gene will be present, a is the average size of the DNA fragments inserted into the vector, and b is the total size of the genome.

Table 6.1 shows the number of clones needed for genomic libraries of a variety of organisms, constructed using a λ replacement vector or a cosmid. For humans and other mammals, several hundred thousand clones are required. It is by no means impossible to obtain several hundred thousand clones, and the methods used to identify a clone carrying a desired gene (Chapter 8) can be adapted to handle such large numbers, so genomic libraries of these sizes are by no means unreasonable. However, ways of reducing the number of clones needed for mammalian genomic libraries are continually being sought.

One solution is to develop new cloning vectors able to handle longer DNA inserts. The most popular of these vectors are **bacterial artificial chromosomes (BACs)**, which are based on the F plasmid (p. 16). The F plasmid is relatively large and vectors derived from it have a higher capacity than normal plasmid vectors. BACs can handle DNA inserts up to 300 kb in size, reducing the size of the human genomic library to just 30,000 clones. Other high-capacity vectors have been constructed from bacteriophage **P1**, which has the advantage over λ of being able to squeeze 110 kb of DNA into its

capsid structure. Cosmid-type vectors based on P1 have been designed and used to clone DNA fragments ranging in size from 75 to 100 kb. Vectors that combine the features of P1 vectors and BACs, called **P1-derived artificial chromosomes (PACs)**, also have a capacity of up to 300 kb.

6.5 Vectors for other bacteria

Cloning vectors have also been developed for several other species of bacteria, including *Streptomyces*, *Bacillus*, and *Pseudomonas*. Some of these vectors are based on plasmids specific to the host organism, and some on **broad host range plasmids** able to replicate in a variety of bacterial hosts. A few are derived from bacteriophages specific to these organisms. Antibiotic resistance genes are generally used as the selectable markers. Most of these vectors are very similar to *E. coli* vectors in terms of their general purposes and uses.

Further reading

FURTHER READING

- Bolivar, F., Rodriguez, R.L., Green, P.J. et al. (1977) Construction and characterization of new cloning vectors. II. A multipurpose cloning system. *Gene*, 2, 95–113. [pBR322.]
- Frischauf, A.-M., Lehrach, H., Poustka, A. & Murray, N. (1983) Lambda replacement vectors carrying polylinker sequences. *Journal of Molecular Biology*, 170, 827–842. [The λ EMBL vectors.]
- Iouannou, P.A., Amemiya, C.T., Garnes, J. et al. (1994) P1-derived vector for the propagation of large human DNA fragments. *Nature Genetics*, 6, 84–89.
- Melton, D.A., Krieg, P.A., Rebagliati, M.R., Maniatis, T., Zinn, K. & Green, M.R. (1984) Efficient *in vitro* synthesis of biologically active RNA and RNA hybridization probes from plasmids containing a bacteriophage SP6 promoter. *Nucleic Acids Research*, 12, 7035–7056. [RNA synthesis from DNA cloned in a plasmid such as pGEM3Z.]
- Sanger, F., Coulson, A.R., Barrell, B.G. et al. (1980) Cloning in single-stranded bacteriophage as an aid to rapid DNA sequencing. *Journal of Molecular Biology*, 143, 161–178. [M13 vectors.]
- Shiyuza, H., Birren, B., Kim, U.J. et al. (1992) Cloning and stable maintenance of 300 kilobase-pair fragments of human DNA in *Escherichia coli* using an F-factor-based vector. *Proceedings of the National Academy of Sciences of the USA*, 89, 8794–8797. [The first description of a BAC.]
- Sternberg, N. (1992) Cloning high molecular weight DNA fragments by the bacteriophage P1 system. *Trends in Genetics*, 8, 11–16.
- Yanisch-Perron, C., Vieira, J. & Messing, J. (1985) Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene*, 33, 103–119.