MICROBIAL GENETICS

Plasmids

Prof. Arif Ali
Professor & Coordinator
Biotechnology Program
Jamia Millia Islamia (A Central University)
Maulana Mohammad Ali Johar Marg
Jamia Nagar, New Delhi - 1100 25, India
E-mail: aliarif@rediffmail.com

02-Dec-2006 (Revised 15-Nov-2007)

CONTENTS
Introduction
General Account
Naming plasmids
Plasmids may be circular or linear
Size of plasmids
Plasmids copy number
Setting the copy number
Plasmid classification
Plasmid incompatibility
Replication, transfer and recombination in plasmids
Insertion sequence (IS elements)
Transposons and controlling elements
Retroelements
Application of plasmids

Key words
de novo, exonucleases, simulations, pars sequence, Cosmids, relapsing fever, viroids, encapsulates, colicin, pathogenicity, pili, phagemids, invertrons, complementation, polar mutations.
Introduction

Bacteria are classically described as haploid organism, carrying their genome on a generally circular chromosome. The indication for the presence of extrachromosomal elements or plasmids became available from the early work on bacterial genetics. They were subsequently studied and described in detail from different bacteria. Lederberg’s group and also that of Heyes while studying sexual reproduction in bacterium in *E. coli* demonstrated that the inheritance of fertility (*F*') phenotype has all the characteristics of an extrachromosomal mode and proposed that a sex factor controls this phenotype independent of the chromosome. Subsequent studies have shown that

(i) Plasmids are DNA molecules, but are not indispensable, and therefore, may or may not be present in a cell;

(ii) When absent, episomes cannot originate *de novo* and can be acquired only from other strains through different modes of genetic exchanges;

(iii) When present, plasmids may be found in autonomous (free) or integrated state (integrated in to chromosomes). In other words they may change their state from autonomous to integrated or vice-versa. Such plasmids (as also some temperate phages) have been referred as episomes, a term which is not popularly used;

(iv) When present, plasmids may be lost irreversibly, also referred as plasmid curing.

With the help of above criteria, one may include sex factors (F), resistance factor ® and colicinogenic factors (Col) in the class of genetic elements called plasmids. The list of plasmids described from bacteria is a long one and ever expanding. One may visualize that, whenever present, these genetic elements are called plasmids and form an accessory genetic system found in addition to main bacterial chromosomes. Additional terms such as IS elements; transposons and retroelements have recently become popular to describe some of the accessory genetic elements found both in prokaryotes and eukaryotes.

General account

Plasmids are pieces of DNA that exist separate from the chromosome. They contain an origin of DNA replication and thus, replicate independent of the chromosome. Plasmids can be as small ranging from a few thousand base pairs to hundred thousand base pairs. Most plasmids are small, from about 0.2 to 4% the size of the bacterial chromosome (In *E. coli* the chromosome is 4639 kilobase pairs). There are families of plasmids whose members have very similar sequences. Plasmids are faithfully transmitted to daughter cells to ensure that they are stably maintained in the population. In addition to an origin for DNA replication, plasmids contain a variety of other genes. These other genes include antibiotic resistant determinant, genes that allow the cell to use a variety of different carbon sources, or genes that are involved in causing diseases, to name a few (Fig. 1a & 1b).

Naming plasmids

Historically, a plasmid has been named to give some indication as to the function carried by the plasmid. Hence we have the F factor (fertility factor), which allow transfer of cell’s chromosomal DNA from one cell to another, the R factor (resistance factor), which carries a large number of antibiotic resistance genes and ColE1, which produce the antibiotic colicin. For
historical reasons, these names have persisted. The convention now is to name a plasmids beginning with the small p, followed by the designation unique to that plasmid. For example, pBR322, the small p indicates that it is a plasmid, B and R are the initials of the persons (Boliver & Rodriguez) who constructed the plasmid and 322 is unique designation number of the plasmids. Naming of many plasmids is still based on the unique function they carry, such as p sym (carrying the symbiotic nitrogen fixing genes), and p Toc (gene responsible for toluene degradation) etc.

Fig. 1a: Bacterial DNA showing chromosomal and Plasmid DNA

Fig. 1b: Plasmid DNA replication and transmission
J. Lederberg (1951) originally used the word plasmid for any extrachromosomal hereditary determinant. This term is now used in a more restricted sense for those accessory DNA circles, which are found in bacteria (or in cell organelles like mitochondria in eukaryotes), in addition to main chromosome. Since in bacteria the main genome is considered to be the chromosome, plasmids can be regarded as an extranuclear genetic material capable of autonomous replication. Therefore, plasmids are included in a treatment of extranuclear genetic systems, which mainly dealt with cell organelles like chloroplasts and mitochondria. However, in eukaryotes, a cell is divided into nucleus and cytoplasm and the organelles are located in the cytoplasm, a parallelism between the cytoplasmic organelles and plasmids can only be established if one considers the organellar genome. Plasmids have, however been studied most extensively using modern techniques of molecular genetics and, therefore deserve a separate detailed treatment.

Plasmids differ among themselves in their characteristics and share the following general properties:

(i) They are genetic elements made up of DNA.
(ii) They are smaller than and separate from the main chromosome and,
(iii) They are capable of autonomous replication.

There are other properties that are characteristic of a plasmid like sex factor, that promote bacterial conjugation.

Phages, which are included in the category of episomes, are often excluded from the categories of plasmids. However, some phages like lambda (\(\lambda\)) share a few properties of plasmids and are therefore, grouped with them by some workers. Phages are, however, able to multiply within the bacterial cell leading to lysis.

**Plasmids may be circular or linear**

The first plasmid described was circular double stranded DNA molecules. In fact, all of the plasmids in the most widely studied bacterium, *E. coli*, are circular. However, circular plasmids are not the only possibility. Once people began looking for linear molecules, more and more linear plasmids and chromosomes were found. Linear plasmids have been characterized from *Streptomyces* species and from the bacterium that causes Lyme disease, *Borrelia burgdorferi*. These bacteria also contain linear chromosomes. Linear chromosomes or plasmids pose several specific challenges (Fig. 2). First, double stranded DNA ends are the substrate for many exonucleases, and as such are very unstable in cells. The linear plasmids must, therefore, protect their ends. Second, double-stranded DNA cannot be replicated all the way to the end of the molecule because of DNA polymerase’s requirement for a primer. In *Borrelia*, this problem has been solved by the plasmids having covalently closed hairpins at the ends to protect them (Fig. 3). The hairpins are also used to replicate the ends of the plasmids.

**Size of plasmids**

The size of the plasmid varies from ~ 1.0 kb to more than 250 kb (Table 1). Many derived plasmids can be of variable sizes. Size of the plasmids is an important property for gene cloning experiments. Larger plasmids are often less in number compared to smaller ones.
Fig. 2: Challenges posed by linear DNA (hairpin loop structure)

Fig. 3: *Borrelia* plasmids

Table 1: Describing size of some plasmids

<table>
<thead>
<tr>
<th>Plasmids</th>
<th>Size (Nucleotide length in kb)</th>
<th>Organism</th>
</tr>
</thead>
<tbody>
<tr>
<td>pBR322</td>
<td>4.363</td>
<td><em>E. coli</em></td>
</tr>
<tr>
<td>ColE1</td>
<td>6.36</td>
<td><em>E. coli</em></td>
</tr>
<tr>
<td>F</td>
<td>95</td>
<td><em>E. coli</em></td>
</tr>
<tr>
<td>TOL</td>
<td>117</td>
<td><em>Pseudomonas putida</em></td>
</tr>
<tr>
<td>pTiAch5</td>
<td>213</td>
<td><em>Agrobacterium tumefaciens</em></td>
</tr>
</tbody>
</table>
Plasmid copy number

A plasmid is present in a cell in a defined number of copies, controlled by its origin of replication. Plasmids such as F or phage- P1 plasmid are present in one copy per cell. Others such as pSC101 are present in 10-15 copies per cell. These are considered low copy number plasmids. The ColE1 plasmids are high copy number plasmids found in approximately 50 copies per cell. Other plasmids have been specifically engineered to be present in extremely high copy numbers.

When a cell divides, the plasmid molecules are inherited by both daughter cells. For the high copy number plasmids, there are enough plasmid molecules for each daughter cell to receive at least some (Fig.4a). The daughter cells do not have to inherit exactly half of the plasmid molecules. Thus, even if a daughter cell only receives a fewer copies of the plasmid than the copy number, this can be redeemed through replication of the plasmid. For this reason, high copy number plasmids do not have to have a strict control to segregate or partition plasmid molecules into daughter cells and are often referred to as “relaxed”. Another problem faced by the plasmid during partitioning is the multi-number formation between plasmid copies. This hinders the plasmid segregation. ColE1 is an example of a high copy plasmid that contains a function to aid in stability of the plasmid. Its cer gene encodes a product that converts plasmid multimers into plasmid monomers (Fig.4b). This ensures that all copies of the plasmid partition independently. While Cer increases the stability of ColE1 plasmids, it is not essential for maintenance of ColE1 in a population.

For the low copy number plasmids, partition of plasmid molecules into daughter cells is much more critical. Because there are very few plasmid copies, if one daughter inherits two molecules, the other daughter will not inherit any plasmid molecules (Fig.4c). Thus, the low copy number plasmids must have mechanisms to ensure their proper partitioning (Fig.4d). P1 has a specific site in the plasmid, called parS, which is required for segregation. The ParA and ParB proteins, which are also produced by the plasmid, bind to the parS sequence. It is thought that one function of the parS-parA-parB complex is to align all the of the plasmid molecules together in the middle of the cell until division is initiated. Then the plasmid molecules are partitioned into daughter cells by the combined action of Par A/ Par B. The result of this partition system is that the low copy number P1 plasmid is appropriately segregated into the daughter cells.

Computer simulations of random segregation of molecules in dividing cells have shown that a partition mechanism is required for plasmid stability, once the copy number of the plasmid is approximately 5-10 per cell. Plasmids with copy numbers above this will be stably maintained in the population even by random segregation. Plasmids with a copy number below this must have a specific mechanism for plasmid partitioning or they may be rapidly lost from the population.

Setting the copy number

Introducing either a low or high copy number plasmid into cells devoid of any plasmid has shown that the copy number is usually fixed by controlling how often replication could initiate. If a low copy number plasmid such as pSC101 is introduced the plasmid replicates only enough so that daughter cells inherit a few copies of the plasmid. If a ColE1 plasmid is introduced, the
plasmid undergoes many more rounds of DNA replication so that there are enough copies of the plasmid for each daughter cell to inherit approximately 50 plasmid molecules.

Fig. 4: Low and high copy number of plasmids

Several general strategies for maintaining copy number have been described for plasmids isolated from E.coli (Fig.5) both strategies employ negative regulation tactics. In one group of plasmids, the amount of initiator protein that is produced is regulated by a specific inhibitor. In a second strategy, multiple binding sites for the initiator protein compete with the origin for binding of a limited pool of initiator protein. In a third strategy, an RNA molecule that a
required for initiation is prevented from binding to the origin by a second RNA and a protein. These strategies regulate the frequency of initiation of DNA replication.

**Fig. 5: Strategies for maintaining copy number of a plasmid**

**Plasmid classification**
There are number of ways by which plasmids can be classified. One such scheme is being discussed below:

**(a) Classification by Phenotype**
The simplest but least useful criterion is the phenotype they confer upon the host cell. The phenotype conferred by a bacterial plasmid does not reflect any intrinsic property of the plasmid molecule itself; and several distinct phenotypes can be conferred by the same plasmid. Mega plasmids carry genes for resistance to many antibiotics and are of great importance to developing therapeutic drugs. It is therefore, difficult to devise a methodical system of nomenclature for
plasmids based on phenotype alone. Many eukaryotic plasmids lack a phenotype, therefore, in such cases phenotype can not act as basis for their classification.

(b) Classification by Structure

Although most plasmids exist as double stranded closed circles of DNA the definition of a plasmid does not exclude other structures. A number of single stranded circular DNA plasmids have been identified in *Streptomyces* and *Clostridium* species, and linear double stranded DNA plasmid have been isolated from several bacterial and Eukaryotic sources, e.g. linear plasmids in *Borrelia hermsii* encode variant surface antigens and are responsible for relapsing fever. In eukaryotes, with linear plasmids, it is not always clear which elements should be described as plasmids and which as chromosomes (i.e., which elements are extra chromosomal and which are part of the genome). In essence, there is no structural distinction between plasmids and chromosomes. All the linear replicons belong to a class of genetic element called invertrons, which have inverted terminal repeats (Irs) with their 5’ ends covalently linked to a terminal protein. Information pertaining to the genetic organization of linear plasmids with inverttron structures is limited, indeed the only linear plasmid pSCL1 from *Streptomyces clavuligerus* has been completely sequences.

There are also RNA plasmids; Viroids are infectious agents composed exclusively of a single piece of circular single stranded RNA that has some double-stranded regions. Viroids mainly cause plant diseases but have recently been reported to cause a human disease.

![Graphic representation of a typical viroid showing some area of double stranded sequences](image)

Fig.: Graphic representation of a typical viroid showing some area of double stranded sequences

Some bipartite, linear, double stranded RNA elements found in yeast also confirmed the definition of a plasmid. These are known as killer factors because they confer a killer phenotype upon the host. They encode a coat protein that encapsulates the genome rather like that of an RNA virus but cannot infect other cells and are transmitted intracellular. The killer factors thus occupy a middle ground between a plasmid and a virus and can be classified as sub-viral agents. It is to be noted that killer factors are not the same as killer plasmids; the latter are more conventional yeast DNA plasmids that also confer a killer phenotype upon the host cell.

(c) Classification by copy number

Plasmids can be categorized on the basis of number of copies present per cell as,

1. High copy number plasmids, also referred as “relaxed plasmids” – Which are normally maintained at multiple copies per cell.
2. Low copy number plasmids or “Stringent plasmids” – Which have a limited number of copies per cell and the same is strictly regulated.
Plasmids with larger copy number are more useful for gene cloning experiments.

**(d) Classification on the basis of intrinsic properties**

A better plasmid classification system uses intrinsic properties such as transfer, replication, maintenance mechanisms, and drug resistance and colicin production. In bacteria, plasmid transfer occurs through transformation and conjugation. Often it is useful to transfer a non-transmissible plasmid to a specific host cell. It is possible to transfer the purified DNA as long as a genetic selection is available for recipients that possess the plasmid. Uptake of purified DNA is called transformation. Some species of bacteria are naturally transformable. Other bacteria that are most useful for genetic engineering are not naturally transformable. In these cases the bacteria require chemical transformation or electro-transformation. Conjugation is an active process, and many plasmids carry gene that promote self-transfer between cells by this method. Bacterial plasmids may thus be classified into two major groups, conjugative and non-conjugative depending upon whether or not they carry a set of transfer gene called the *tra* genes. The *tra* consists of a large (~33 kb) region of the F plasmid, called the transfer region, it contains ~40 genes; these genes are named as *tra* and *trb* loci, and the majority are expressed coordinately as part of a single 32 kb transcription unit. These *tra* genes promote bacterial conjugation. Generally conjugative plasmids are of high molecular weight and are present as 1-3 copies per cell, whereas non-conjugative plasmids have low molecular weight and are present in multiple copies i.e. 20-25 copies per cell.

Based on the intrinsic properties the three main types of plasmids are:

1. **F or sex factor** - responsible for transfer of genetic material from one strain to another
2. **R factor** - responsible for drug resistance
3. **Col factor** - responsible for colicin production.

Some plasmids may share the properties of more than one type of plasmids but they are still classified only in one of the three categories. For instance, some R plasmids control drug resistance as well as transmissibility. There may be other plasmids, which have properties other than the three described above. For instance, some plasmids confer pathogenicity on host bacteria. This classification therefore, has been considered unsatisfactory, but no other satisfactory classification has been evolved so far. The three types of plasmids, F, R and Col. will be briefly described in this section Table 2.

*The F factor or sex factor*

When F\(^+\) strain (carrying F factor) comes in contact with F\(^-\) two processes may follow, one involving transfer of F factor, so that F\(^-\) becomes F\(^+\) and second involving transfer of a segment of bacterial chromosome. The latter is facilitated by insertion of F factor within the bacterial chromosome at one of the several specific sites available, leading to the production of Hfr (high frequency recombinant) strain.

F factor has been included in the category of cell constituents called episomes, since F can exist in ‘free’ or ‘integrated’ state. However all plasmids are not episomes, since they are not always able to integrate in to the main chromosome. Nevertheless F is a plasmid, because it is an
autonomous DNA molecule, smaller than main chromosome and can replicate independently of bacterial chromosome.

**Table 2: Characteristics of some bacterial plasmids**

<table>
<thead>
<tr>
<th>S. No</th>
<th>Plasmid</th>
<th>Characteristic</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>F, RI, Col</td>
<td>Fertility/antibiotic resistance/colicin production</td>
</tr>
<tr>
<td>2</td>
<td>ColE1</td>
<td>Bacteriocin production</td>
</tr>
<tr>
<td>3</td>
<td>R6</td>
<td>Heavy metal resistance</td>
</tr>
<tr>
<td>4</td>
<td>Ent</td>
<td>Enterotoxin production</td>
</tr>
<tr>
<td>5</td>
<td>Cam</td>
<td>Metabolism of camphor</td>
</tr>
<tr>
<td>6</td>
<td>Ti (in <em>Agrobacterium tumefaciens</em>)</td>
<td>Tumorigenicity in plants</td>
</tr>
</tbody>
</table>

*The R plasmid or resistance transfer factor (RTF)*

R plasmid or R factor, when present, confers on bacteria, resistance to several antibiotics, such as streptomycin, tetracycline, chloramphenicol and sulphonamide. This was discovered in Japan in 1956, when a bacterial strain called *Shigella* was shown to acquire resistance against several drugs in one step. This was shown to be due to transmission of multiple resistances from one strain to another and was later attributed to R plasmid. R plasmid, not only confers resistance, but also enables bacteria carrying them to conjugate and transfer it to other bacteria, related or quite unrelated. The genes for many drug resistances were later shown to be due to transposons.

It has been shown, that like F factor, R is also a small extrachromosomal DNA circle, and is autonomous. Sometimes a plasmid may consist of two parts, one bearing genes for resistance (R factor) and the other responsible for transfer of plasmid (resistance transfer factor or RTF). In some cases, like penicillin resistance in *Staphylococcus*, the resistance may not be associated with ability to transfer (RTF).

*The Col factors*

Colicins are toxic proteins which are produced by bacteria (*Escherichia, Shigella* and *Salmonella*) and which kill bacteria other than those that produce them. This property of bacteria is genetic in origin and depends on the elements called colicinogenic factors, such plasmids now are popularly known as Col plasmids. They are one class of a general type of plasmid called bacteriocinogenic plasmids, which produce bacteriocins in many bacterial species. Bacteriocins, of which colicins are one example, are proteins that can interact with sensitive bacteria and inhibit one or more essential processes such as DNA replication, transcription, translation, or energy metabolism.

Col factors resemble phages, since they cause death of bacteria; however, they differ from phages, in the sense that the death of bacteria is not accompanied with simultaneous release of Col factors as in the case of lysis caused due to phages.
Genetic studies suggest that colicin production is not determined by genes located on bacterial chromosomes but by genes present on Col plasmids, which can be transmitted from one bacterium to another. The best-studied Col plasmid is ColE1. It is a mobilizable but non-conjugative plasmid. The transfer of ColE1 requires a plasmid-encoded mob gene; it codes for an enzyme called nuclease, and a specific base sequence called bom (basis of mobility), which contains the cutting site. In a critical experiment it was shown that the compatible plasmid F and ColE1 were present in the same cell, and F donates the conjugal functions that ColE1 lacks, thereby enabling ColE1 to be transferred.

(e) Degradative Plasmids

Many species of *Pseudomonas* can use different organic compounds as carbon sources – in particular, toxic organic compounds like camphor, toluene (TOL plasmid), octane, and others allow the host bacterium to metabolize unusual molecule such as salicylic acid. Each plasmid provides one or more metabolic pathways to degrade these compounds. Because many enzymes are needed for this purpose, the plasmids are fairly large. Some strains can degrade persistent chlorinated hydrocarbons, herbicides, pesticides and various detergents. Plasmids have also been isolated that confer resistance to heavy metal ions. These plasmids are found in environment containing these ions, such as the sludge produced from industrial reprocessing of photographic film (resistant to the Ag\(^+\) ions). The biochemistry of resistance to Hg\(^{2+}\) ions as well as many other toxic metal ions has been fairly well understood: The resistance results from plasmid-encoded genes to detoxify the metals.

(f) Virulence plasmids

These confer pathogenicity on the host bacterium, e.g. tumor causing ability of Ti plasmids of *Agrobacterium tumifaciens* plants, which induce crown gall disease on dicotyledonous plants. When a plant is infected, some of the bacteria enter and grow within the plant cells and lyse, releasing their DNA in the cell. From this point on, the bacteria are no longer necessary for tumor formation. Small segment of the Ti plasmid (the T-DNA), containing the genes for replication, become integrated into the plant cell chromosomes. The integrated fragment modulates the hormonally regulated system that controls cell division, causing the cell to be converted into a tumor cell.

(g) Other plasmids

Several plasmids render fairly innocuous plasmids-free bacteria pathogenic. The Ent plasmids of *E. coli* synthesize enterotoxins that are responsible for travellers’ diarrhea. A plasmid called Hly (for hemolysis) has been found in *E. coli* strains isolated from pigs. These plasmids destroy red blood cells in blood samples but do not seem to cause any pathogenicity.

In eukaryotes, plasmid transfer mechanism is a less useful criterion for classification. Horizontal plasmid transfer generally occurs only when cells fuse (e.g. syngamy, or the formation of hyphal network in fungi) or occasionally by mechanical transfer (viroids spread in this manner). Occasionally, bacteria transfer plasmids to eukaryotes, as occurs in bacteria to yeast conjugation and in the specialized case of the *Agrobacterium tumifaciens* Ti plasmid.
**Plasmid incompatibility**

Cells can maintain more than one plasmid at a time, faithfully transmitting multiple plasmids to daughter cells. Maintenance of more than one plasmid in a cell can only occur if the plasmids carry different origins of replication in terms of their replicational requirements. The inability of two plasmids with the similar origin to be maintained in the same cell is known as incompatibility. Incompatibility is thought to be related to limiting concentrations of the initiator proteins and how cells pick which plasmid molecules are replicated (Fig.6). When the concentration of initiator protein is high enough, one plasmid molecule from the population is randomly chosen and replicated. Both of the newly replicated plasmid molecules then become part of the pool of plasmid molecules and this continues till the initiator protein level becomes limiting. If two different plasmids that contain the origin of replication requiring the same initiator protein, are introduced into the same cell, one is lost after a few generations precisely because of this random picking of plasmid molecules for replication. If there is one copy of plasmid A and one of plasmid B either plasmid A or B can be chosen to be replicated. If plasmid A is replicated then the cell will contain two molecules of plasmid A and one of plasmid B. This means that when the next plasmid molecule is chosen at random for replication, there is a greater probability that it will be a molecule of plasmid A. In a few cycles, plasmid B will be lost from the cell and plasmid A maintained. Incompatibility is not only important for the maintenance of plasmids in a population of cells but also in cloning experiments where plasmids are used to carry specific pieces of chromosomal DNA.

![Plasmid incompatibility diagram](source)

**Fig. 6: Plasmid incompatibility**

*(Source: Nancy Trun & Janine Trempy. Fundamental bacterial genetics, 2004)*
Replication, transfer and recombination in plasmids

A plasmid must contain an origin for DNA replication (ori). In the absence of an ori site, the plasmid cannot replicate and when the cells divide, one daughter cell would inherit a plasmid molecule and the other would not. At each subsequent cell division, only one daughter cell would receive a plasmid molecule (Fig. 7). The scenario arising when a plasmid lacks an origin of replication is the loss of plasmid.

If *E. coli* are inoculated into growth media and incubated overnight, they will divide approximately 20 times. If one cell in the population contained one plasmid molecule without an origin of replication, at the end of this period of cell growth, one cell would have a plasmid molecule and 1,048,575 cells would not! Without an origin of replication, a plasmid is lost from the cell population very rapidly.

An origin consists of a small stretch of DNA that binds specific proteins and/or RNA molecules (Fig. 8a). These components must open the double helix of the DNA at or near the origin sequence, provide for the synthesis of a primer for DNA polymerase, and provide for the continued DNA synthesis of each strand of the plasmid DNA molecule. As can be imagined, not every origin of replication functions identically. There are many different ways to accomplish the functions that an origin of replication is responsible for. Each plasmid has its own scheme for replication and many variations have been described.

Several plasmids encode a protein required for initiation of DNA replication and other do not (Fig. 8b). This plasmid-encoded initiator protein binds to the ori region and helps to form an

Fig. 7: Consequences of a plasmid not having an origin of replication
open complex. Frequently, the host-encoded protein DnaA binds to the origin along with the plasmid-encoded initiator protein to facilitate this reaction. In other plasmids, the plasmid encodes several proteins required for initiation of replication and DnaA is not required (Fig. 9). Still other plasmids use an RNA molecule to make an open complex at the ori (Fig.10). This RNA is also used as a primer for DNA polymerase.

![Diagram of DNA replication](https://via.placeholder.com/150)

**Fig. 8: Two examples of the structure of an origin of DNA replication (ori).**

**Fig. 9: Some plasmids encode all of their own initiator proteins**

DNA replication process is carried out using several different mechanisms:
- DNA synthesis from the origin proceeding in one direction at a time (unidirectional DNA synthesis) or in both directions simultaneously (bidirectional DNA synthesis) (Fig.11).
- Rolling circle replication

Plasmids also differ in their requirement for different host proteins as well as for RNA polymerase for initiation of the replication. Because plasmids have a mechanism to replicate their DNA and ensure that they are stably maintained in a cell population, they are also known as
replicons. Any DNA molecule that has these two properties, including the cell’s chromosome, is a replicon.

![Diagram of ColE1 plasmid replication](image)

**Fig. 10: ColE1 plasmids use an RNA molecule to form the open complex of ori and initiate DNA synthesis**


All plasmids identified so far are small circular DNA molecules. The DNA circles may be single or multiple structures linked together in chains. The length of these plasmid DNA molecules varies from 30 \( \mu \)m of *E. coli*, making 20% of the main chromosome (same size in R plasmids also). Colicin plasmids ColE\(_1\), E\(_2\) and E\(_3\) are usually small having a length of 2.3 \( \mu \)m. The number of plasmids in a cell varies from one to a few in case of F and R plasmids to as many as 20-30 in case of *Staphylococcus* plasmids, and 50-100 in case of pUC plasmids.

DNA plasmids replicate in semi-conservative manner. It has been shown that while initiation of replication is controlled by plasmid genes, other stages of replication i.e., polymerization
(elongation) and termination are controlled by host genes. When a plasmid is inserted in the bacterial chromosome, bacterial control on plasmid replication is absolute, but when free, the control is less that absolute.

Fig. 11: Plasmids can replicate unidirectionally or bidirectionally

Plasmids also have the ability of transferring themselves from one bacterium to another. This transfer is facilitated by the formation appendage known as sex pilus, extending out of the outer membrane of the donor bacterium (refer bacterial conjugation). The F sex pilus is up to 20µm in length and 8 nm wide. The sex pili produced by different plasmids are distinguished by the response they show for attachment of certain phages. When sex pilus is mechanically removed
or is absent due to a genetic mutation, no transfer of plasmid DNA is possible, suggesting that sex pilus is essential for transfer.

In *E. coli*, transfer is controlled by plasmid genes known as *tra*, several of which are known, -A, -B, -C, etc., and mutations in these genes lead to failure of transfer. However, when two plasmids having mutations for different *tra*-genes are introduced in the same cell, pilus may be formed and transfer may occur due to complementation between mutants. When transfer occurs, one of the two strands of plasmid DNA is nicked (broken) at a site called origin of transfer (*OriT*) and the linear strand thus formed moves to the recipient bacteria. The single strands in the donor as well as in recipient cell will then synthesize complementary strands, so that duplex DNA plasmids are reconstituted in both the cells (Fig. 12). Sometimes due to insertion of plasmid, the transfer process may involve transfer of bacterial genes also, which are carried along with plasmid DNA.

![Fig. 12: Five different stages of infection transfer](Source: Nancy Trun & Janine Trempy. Fundamental bacterial genetics, 2004)

Plasmid DNA may undergo recombination either with another plasmid DNA or with in the bacterial chromosome. In both, there are insertion sequence (IS) which is involved in homologous recombination. There are at least 17 such sites in *E. coli* chromosome. Mapping of genes in plasmid DNA has also been achieved. Methods employed in mapping are not only genetic but also physical.

**Insertion sequences or IS elements**

In prokaryotic as well as eukaryotic genomes, certain sequences are capable of moving from one site to another and are described as transposable elements (TE). Prokaryotic TEs may include (i) insertion sequences or IS elements. (ii) transposons, also characterized by certain phenotypes other than their ability to transpose and (iii) retroelements, which involve reverse transcription, either in their origin or propagation.

Insertion sequences or IS elements were the first transposable elements, identified as spontaneous insertion in some bacterial operons. These insertions inactivate the gene and do not allow transcription and translation of this and the following (other genes downstream) genes in the operon. The first operon, where IS elements were detected, was *gal* operon of *E. coli* (Fig.13) responsible for synthesis of three enzymes (epimerase, transferase, galactokinase) needed for
metabolism of galactose. These mutations caused by IS elements were called polar mutations, since transcription from all sequences downstream from insertion site was inhibited. Since these mutations did not revert using different mutagens, and because different mutagens did not affect frequency of reversion, these were neither deletions, nor frameshift or point mutations. In order to study the mechanism of induction of polar mutations, lambda (\(\lambda\)) phage (which picks up gal region after inserting in this region and excising out) was used. \(\lambda\,dgalm\) was isolated from mutant \(gal^m\,E.coli\) genome, and its DNA was used for heteroduplex analysis along with \(\lambda^+\) DNA. For this, the two DNAs were denatured and then renatured. This DNA was visualized by electron microscopy where a heteroduplex (\(\lambda_{gal^m}\lambda^+\)) was seen to carry an extra piece of DNA (768 bp) and could be detected in the form of a loop (Fig. 14). This was designated as IS1. Several such IS elements were later detected in \(E.coli\) (Table-3) when many other polar mutants were analyzed identifying different IS elements. All of them are characterized by:

(i) a characteristic size;
(ii) the presence of inverted terminal repeats (TR) required for transposition;
(iii) the ability to generate, at the target site, direct repeats of flanking DNA; and
(iv) presence of open reading frames, coding for enzyme transposase, which is essential for transposition.

**Fig. 13: The gal operon of \(E.coli\)**
(Source: Benjamin Lewin-Genes VI, 1997)

**Fig. 14: \(\alpha\,dal^+\alpha\,dgal^m\) DNA heteroduplex’s electron micrograph**
Table 3: Some prokaryotic insertion sequence (IS) elements
(Source: Benjamin Lewin-Genes VI, 1997)

<table>
<thead>
<tr>
<th>Element</th>
<th>No. of copies Per genome</th>
<th>Length (bp)</th>
<th>Inverted TR (bp)</th>
<th>Direct Repeat at target (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>IS1</td>
<td>5-8</td>
<td>768</td>
<td>23</td>
<td>9</td>
</tr>
<tr>
<td>IS2</td>
<td>6</td>
<td>1327</td>
<td>41</td>
<td>5</td>
</tr>
<tr>
<td>IS3</td>
<td>7</td>
<td>1400(app)</td>
<td>38</td>
<td>_</td>
</tr>
<tr>
<td>IS4</td>
<td>1 or 2</td>
<td>1428</td>
<td>18</td>
<td>11-12</td>
</tr>
<tr>
<td>IS5</td>
<td>unknown</td>
<td>1195</td>
<td>16</td>
<td>4</td>
</tr>
</tbody>
</table>

Transposons and controlling elements (mediated via DNA)

In the previous section, we described insertion sequences, which are capable of moving within the genome through illegitimate recombination (moving to DNA sequences quite unrelated). In this section, we describe additional extrachromosomal elements, which are transposable and can occupy different sites on the main DNA molecule of plasmid or bacterial chromosome. These elements are popularly described as transposons, both in prokaryotes and eukaryotes. However, in maize these were initially described as controlling elements, due to their control on the expression of genes. The relationship between plasmids, IS elements and transposons (Tn) is shown in Fig.15.

Transposons in Prokaryotes

Transposons (Tn) is a term used in 1974 by R.W. Hedges and A.E. Jacob of Hammersmith Hospital in London, for a DNA segment or genetic element, which could move from one molecule to another and carried resistance for antibiotic ampicillin. They observed that transfer of such antibiotic resistance from one plasmid to another is accompanied by an increase in size of the recipient DNA molecule (or plasmid). This recipient plasmid could donate this resistance to another plasmid, which also showed a similar increase in size thus proving that transfer of a DNA segment was involved. This DNA segment, called a transposon, which can occupy different sites in the genome, can be transposed between a plasmid or from plasmid to bacterial chromosome or vice versa. It was also shown that transfer of transposon carrying ampicillin resistance could take place not only in plasmids as shown in Fig.16, but even in those bacteria which were mutant for the gene recA responsible for recombination, thus suggesting that these
transfers did not involve normal recombination process. Properties of some of the known transposons are given in Table 4.

It was also shown that two ends of each of the two strands of transposon consisted of nucleotide sequences that were complementary to each other, but in reverse order. For instance if one end had CCCAGAC, the order end will have GTCTGGG. These inverted repeats will form stem and loop structure characteristic of such repeats. When a plasmid with transposon was denatured and each single strand was allowed to base pair among its ownself, such stem and loop structures of different sizes were actually observed depending upon the Tn type (Fig.17). All transposons studied so far have been found to consist of inverted repeats at the two ends. These repeated sequences may range from few basepairs to as many as 1500. It has been shown that insertion of any gene between two IS elements could allow the transfer of such a gene possible without normal recombination process. Therefore, it has been observed that genes, other than those for antibiotic resistance, which are known to be generally absent in transposons, have been found to accompany the transposition. It has also been shown that some transposons have their ends consisting of insertion sequences (IS) which themselves resemble a TE in having inverted repeats and the ability to move. This means that two identical IS elements with an intervening sequence may give rise to a new composite transposon (Fig. 18).

**Fig. 16: Transposition of transposon carrying Amp from one plasmid**
Table 4: Properties of some well known prokaryotic transposons

<table>
<thead>
<tr>
<th>Transposons</th>
<th>Resistance against</th>
<th>Length (bp)</th>
<th>Inverted repeat (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tn1, Tn2, Tn3</td>
<td>Ampicillin</td>
<td>4,957</td>
<td>38</td>
</tr>
<tr>
<td>Tn4</td>
<td>Ampicillin, Streptomycin,</td>
<td>20,500</td>
<td>short</td>
</tr>
<tr>
<td></td>
<td>Sulphonamide</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tn5</td>
<td>Kanamycin</td>
<td>5,400</td>
<td>1500</td>
</tr>
<tr>
<td>Tn6</td>
<td>Kanamycin, Streptomycin</td>
<td>4,200</td>
<td>unknown</td>
</tr>
<tr>
<td>Tn7</td>
<td>Trimethoprim, Streptomycin</td>
<td>14,000</td>
<td>unknown</td>
</tr>
<tr>
<td>Tn9</td>
<td>Chloramphenicol</td>
<td>2,638</td>
<td>18/23</td>
</tr>
<tr>
<td>Tn10</td>
<td>Tetracycline</td>
<td>9,300</td>
<td>1400</td>
</tr>
<tr>
<td>Tn204</td>
<td>Chloramphenicol, Fusidic acid</td>
<td>2,457</td>
<td>18/23</td>
</tr>
<tr>
<td>Tn402</td>
<td>Trimethoprim</td>
<td>7,500</td>
<td>unknown</td>
</tr>
<tr>
<td>Tn551</td>
<td>Erythromycin</td>
<td>5,200</td>
<td>35</td>
</tr>
<tr>
<td>Tn903</td>
<td>Kanamycin</td>
<td>3,100</td>
<td>1050</td>
</tr>
<tr>
<td>Tn917</td>
<td>Erythromycin</td>
<td>5,100</td>
<td>short</td>
</tr>
<tr>
<td>Tn1721</td>
<td>Tetracyclin</td>
<td>10,900</td>
<td>short</td>
</tr>
</tbody>
</table>

Fig. 17: Presence of inverted repeats at the two ends of each strand of transposon
Retroelements (Transposable elements involving RNA phase)

Retroelements (these are also transposable elements) are those nucleic acid sequences, which either originate partly or fully from an RNA and, or propagate through reverse transcription (RNA—► DNA). These retroelements are classified into (i) Viral retroelements (ii) non-viral retroelements. Both these classes are further subdivided on the basis of structural and biological criteria (Table 5). Since viral retroelements also have an extracellular phase (they are released) and are infectious, they may not be very important for a discussion on extranuclear accessory genomic elements, included in this chapter. The non-viral retroelements include retrotransposons, which resemble transposons, except that they have RNA origin. They have

(i) LTR (long terminal repeat) sequences needed for transposition,
(ii) primer binding site (PBS),
(iii) a promoter, and
(iv) processing/ polyadenylation signal.

These attributes are missing from retroposons. Retrons include msDNA of myxobacterium and E.coli, but may include additional sequences from other prokaryotes. They are characterized by an unconventional mechanism of reverse transcription. Retrosequences include cDNA genes or pseudogenes, which end in poly (A) tails and are flanked by short repeat sequences.

Myxobacteria, a soil bacterium (Stigmatella aurantiaca) and 10% of all clinical strains of E.coli tested have been shown to produce RNA-DNA complexes consisting of (i) a DNA component called multicopy single stranded DNA (msDNA) and (ii) an RNA component called multicopy single stranded DNA like RNA (msdRNA). The genomic DNA consists of (i) sequences coding for reverse transcriptase, (ii) an msd sequence coding for msDNA and (iii) a msr sequence coding for msd RNA. DNA molecule is attached at 5’ end to a G-residue of RNA molecule through a phosphodiester bond, and on the 3’ end it attaches non-covalently via a few residues. Both DNA and RNA have characteristic stem-loop structure (Fig.19).

Application of plasmids

As discussed earlier, plasmid in bacteria provide several functions that have important bearing on bacterial evolution. Most significantly, however, plasmids have become an important tool in recombinant DNA technology as they have provided essential components of the cloning and
expression of a desired gene. Several modern cloning vectors e.g. pBR322, pUC18 and pUC19 have been derived and thus have their ancestry in the wild resident plasmids. There are number of vectors that have been designed for different purpose and a very brief account is been given here.

Table 5: Classification of Retroelements

(A) Viral Retroelements
   (I) Retroviruses (RNA in virions or plus strand viruses) e.g. Oncornaviruses, Lentiviruses, Spumaviruses.
   (II) Pararetroviruses (DNA in virions) e.g., Caulimoviruses, Hepadnaviruses

(B) Non-Viral retroelements
   (I) Retrotansposons
   (II) Retroposons
   (III) Retrons
   (IV) Retrosequence

<table>
<thead>
<tr>
<th></th>
<th>LTR</th>
<th>RT</th>
<th>Integrase</th>
<th>SINEs (Alu)</th>
<th>msDNA</th>
<th>cDNA genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Viral Retroelements</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(I) Retroviruses</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(II) Pararetroviruses</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-Viral retroelements</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(I) Retrotansposons</td>
<td></td>
<td>+</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(II) Retroposons</td>
<td></td>
<td></td>
<td>+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(III) Retrons</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(IV) Retrosequence</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Examples Ty, copia, LINEs

Fig. 19: Structure of msDNA retron of myxobacteria and *E. coli*
**Cloning Vector**

**Expression vectors**

Different types of cloning vectors have been developed to specifically direct the transcription and translation of a cloned gene. Vectors designed for these purposes are collectively known as expression vectors. The promoters used in expression vectors are selected with care. It is most desirable if the cloned gene is expressed under a defined growth condition. To achieve this objective two commonly used promoters, one from the lac operon and one from the ara operon, have been successfully employed. The lac promoter is induced by the addition of lactose (allolactoses) or its non-metabolized analogue called isopropyl-β-D-thiogalactoside (ITPG) to the growth medium. The ara promoter is induced by addition of arabinose. Also a hybrid promoter i.e. tac promoter (lac + trp promoter) induced by ITPG is also being used.

**Shuttle vectors**

These are designed to replicate in two different species. It contains origin of replication and selectable markers recognized in two different species. One of the origins, and one of the selectable markers, functions in one species, and the other origin and selectable marker function in the second species. One of the species generally used where the shuttle vector replicates almost always is *E. coli*. The other species can be as diverse as *Bacillus subtilis*, *Saccharomyces cerevisiae*, rodent cell lines, plants, or human cell lines.

**Vectors for localizing the gene products**

For many proteins, it is desirable to know where in the cell the protein resides. One type of fusion vector that has been developed relies on the green fluorescent protein (GFP) isolated from the jellyfish, *Aequoria victoria*. GFP is a 35-kilodalton protein that accepts energy from blue-green light at an optimum wavelength of 400 nm and emits green light at a wavelength of 509 nm. It means that GFP fluoresces and a microscope equipped with a low light camera and the appropriate filters can detect this fluorescence. A major advantage to GFP is that (Fusion protein) localization can be carried out in live cells.

**Vectors for purifying the cloned gene product**

A group of expression vectors have been designed to drive the expression of as much of the cloned gene product as is possible and use regulated lambda or T7 bacteriophage promoters. The phage promoters are some the strongest known promoters in biological systems. In T7 promoter containing vectors, the cloned gene product can be over expressed until it represents up to 10% of the total cell protein. This much level of the cloned gene product is usually lethal to the cell. Since, the phage promoter vectors are designed and used to purify the gene product, and as long as the cells do not lyse and the objective remains purification of the gene product, it does not matter if the cells are alive or dead.

**Vectors for studying gene expression**

How genes are regulated can be determined by using a vector containing a reporter gene. A reporter gene is a well-characterized gene whose product can be quantitatively measured in an easy assay. The promoter region from a gene of interest is cloned in front of a reporter gene and the amount of reporter gene activity and the condition under which it is produced are examined.
Two types of reporter gene construct can be made. In the first type, the signal needed for transcription (the –35, -10, and +1 site) from the gene of interest is placed in front of the reporter gene. The signals needed for translation (the RBS and starting met codon) are already present in the vector. This type of fusion is known as a transcriptional or operon fusion (Fig 20). In the second type, all of the signals for transcription and translation from the gene of interest are placed in front of the reporter gene. This creates a novel gene whose 5’ end comes from the gene of interest and the 3’ end comes from the reporter gene. This type of fusion is known as a translational or gene fusion (Fig. 21). These types of fusions provide information about how the gene of interest is regulated and whether the regulation is at transcriptional or at translational level. Many of the currently available reporter gene vectors use lacZ. But others like gus, lux or gfp are also common.

**Fig. 20: An operon of transcriptional fusion**

**Fig. 21: A gene of translational fusion**

**Phage vectors**

Many of the features described for plasmid vectors can also be incorporated in phage vectors. Phages are very good at moving DNA into cells. A recombinant phage, after infection, can practically inject every cell with the recombinant DNA molecule, and by isolating phage; a very pure and highly concentrated source of DNA is made available. Some phages are maintained in
the host cell as lysogens in low copy numbers, while other phages produce particles with single stranded DNA in them. Each of these features can be exploited to make a phage vector. Phage vectors have been constructed based on λ, M13, Mu, and P1.

Cosmids are plasmid vectors containing λ phage cos (cohesive end) sites incorporated into the plasmid DNA. Cos sites complement one another and can bind together to circularize the DNA molecule. Cosmids typically contain an origin of replication such as ColEI ori, several unique restriction sites, and one or more selective markers for antibiotic resistance. This permits a cosmid vector to replicate like a plasmid; it can be packaged in λ phage heads and be moved from cell to cell like λ. Wild type λ phage vector could accommodate only 2.5 kb of foreign DNA, whereas Cosmids, the first artificial chromosome vectors, can accept DNA fragments as large as ~40-50 kb (or 10 times the amount of foreign DNA as standard plasmid cloning vectors), and be maintained as stable plasmids. This ability to clone large fragments of DNA is especially useful in cloning eukaryotic genes or large chunk of DNA.

Phagemids are also hybrid vectors containing DNA from a filamentous phage e.g. M13, and a plasmid. These phagemid vectors contain origins of replication for both the phage and the plasmid. Under normal conditions, replication of the phagmid proceeds from the plasmid’s ori site to yield double-stranded DNA molecules. However, when a cell that has already a phagmid in it is infected with a single-stranded filamentous phage (also sometimes called a helper phage), the replication mode changes. One strand of the already existing phagmid double stranded DNA molecule is nicked and a new DNA is synthesized from it by the rolling circle mechanism of replication (refer to book of molecular genetics). In this way, single-stranded copies of the phagmid DNA and along with it the single-stranded copies of the gene cloned in this phagmid are generated; these DNA can be isolated and used for sequencing employing dideoxy chain termination method.

Artificial chromosomes

Most of the vectors developed were designed for inserts of the size of ~10 kilobases of DNA fragment or less. Since, in contemporary times, DNA sequencing of different organisms has become a major research program of different labs all over the world, and with the advent of Bioinformatics to handle the huge DNA sequence data, artificial chromosomes have been constructed that can accept DNA fragments as large as 40 to 1000 kilobases. They are based on the chromosomes of Lambda (cosmids), P1 (PACs), bacteria (BACs), and yeast (YACs) and they are useful in circumventing many problems that the scientists face when involved in sequencing an eukaryotic organism. Artificial chromosomes contain all of the DNA sequences necessary for them to replicate as chromosome. PACs are low copy number plasmids and can take ~90kb; BACs use sequences from the F factor to replicate and can accept several hundred kilobases of DNA and YACs utilize sequences from yeast chromosomes to replicate, and can accept up to 1000 kb of DNA.

Ti plasmid

This plasmid has become important in plant genetic engineering because a part of this plasmid (T-DNA) is transferred and integrated into the plant genome. If specific genes can be tagged into it they can be transferred to a desired plant. New plant varieties having desirable and
economically valuable characteristics derived from unrelated species can be developed in this way.

**Suggested Readings**