Molecular Biology: Gene cloning

Author: Prof Marinda Oosthuizen
Licensed under a Creative Commons Attribution license.

TABLE OF CONTENTS

INTRODUCTION ........................................................................................................................................... 2
GENERAL STEPS OF GENE CLONING ..................................................................................................... 2
CLONING VECTORS ................................................................................................................................... 4
   Plasmids ................................................................................................................................................... 5
   Bacteriophages ........................................................................................................................................ 5
   Cosmids: .................................................................................................................................................. 7
   Yeast artificial chromosomes (YAC): ....................................................................................................... 8
   Bacterial artificial chromosomes (BAC): .................................................................................................. 9
   Expression vectors: .................................................................................................................................. 9
RESTRICTION ENDONUCLEASES .......................................................................................................... 10
LIGATION: THE JOINING OF DNA MOLECULES ................................................................................... 11
INTRODUCTION OF DNA INTO HOST CELLS ........................................................................................ 12
   Transformation: The uptake of DNA by bacterial cells ........................................................................ 12
   Introduction of phage DNA into bacterial cells: ...................................................................................... 14
   Transformation of non-bacterial cells: .................................................................................................... 14
GENE LIBRARIES ...................................................................................................................................... 15
SCREENING AND SELECTION FOR RECOMBINANTS ........................................................................ 15
   Direct selection: ...................................................................................................................................... 16
   Identification of the clone from a gene library: ....................................................................................... 17
ANALYSIS OF A CLONE ........................................................................................................................... 19
INTRODUCTION

What is a clone? In simple terms, a clone is a duplicate, just as a photocopy is a replica or copy of a document. Identical twins, which are duplicates of each other, is an example of such “copies” that occur in nature. These days, one regularly hears about the cloning of animals (sheep, cattle, horses, mice, and even house pets) in the news. From time to time, you may even have heard about researchers cloning, or identifying, genes that are responsible for various medical conditions or traits. But, what is the difference?

*Cloning an animal*, or any other organism, refers to making an exact genetic copy of that organism by using artificial embryo twinning or somatic cell nuclear transfer.

*Cloning a gene* means the isolation of an exact copy of a single gene from the entire genome of an organism. This usually involves copying the DNA sequence of that gene into a smaller, more accessible piece of DNA, such as a plasmid. This makes it easier to study the function of the individual gene in the laboratory.

The terms “recombinant DNA technology”, “DNA cloning”, “molecular cloning” or “gene cloning” all refer to the same process: the transfer of a DNA fragment of interest from one organism to another, and in the process generating multiple copies of the same gene. The word “clone”, therefore, refers to a gene or DNA fragment and also to the collection of cells or organisms (such as bacteria) containing the cloned piece of DNA. Because molecular cloning is such an essential tool of molecular biologists, in scientific circles “to clone” has become synonymous with inserting a new piece of DNA into an existing DNA molecule.

GENERAL STEPS OF GENE CLONING

The general steps of gene cloning with any vector are as follows (Figure 1):

1. The **vector** and the DNA to be cloned (foreign or target DNA) are prepared by **digestion** or “cutting” with restriction enzymes to generate complementary ends.
2. Foreign DNA is **ligated** or “joined” with the vector by using the enzyme DNA ligase.
3. The recombinant DNA molecule is introduced into the host cell (usually bacterial cells) by **transformation** (transfection, transduction, *in vitro* packaging).

The host cells are grown and recombinant colonies (cells containing the foreign DNA) are selected by **screening** for selectable markers (usually antibiotic resistance).
Molecular Biology: Gene cloning

These steps will now be discussed in more detail, but before we go one, some brief definitions to understand the terminology used above:

**DNA ligase:** An enzyme that repairs single-stranded discontinuities in double-stranded DNA molecules in the cell. Purified DNA ligase is used in gene cloning to join DNA molecules together.

**Ligation:** The process of joining two or more DNA fragments together

**Recombinant:** A transformed cell that contains a recombinant DNA molecule.

**Recombinant DNA:** A DNA molecule produced by inserting DNA from one organism (or may be chemically synthesized) into another piece of DNA using gene manipulation techniques.

**Restriction endonucleases:** Bacterial enzymes that cut (hydrolyze) DNA at specific recognition sequences into defined and reproducible fragments. In bacteria, they form part of the restriction-modification defence mechanism against foreign DNA.

**Selection (Screening):** A means of obtaining a clone containing a desired recombinant DNA molecule.

**Transformation:** Transformation is the process of take-up of foreign DNA, normally plasmids, by bacteria.

**Vector:** An agent or vehicle that carries a DNA fragment into a host cell.
CLONING VECTORS

The central component of a gene cloning experiment is the vector or vehicle, which transports the gene into the host cell and is responsible for its replication. If it is used for reproducing the DNA fragment, it is called a cloning vector. If it is used for expressing a certain gene in the DNA fragment, it is called an expression vector. Commonly used vectors include plasmids, bacteriophages (phage Lambda, M13), cosmids, yeast artificial chromosomes (YAC), and bacterial artificial chromosomes (BAC).
Features shared by vectors:

- They are small, well characterized molecules of DNA.
- They contain a replication origin and can be replicated within the appropriate host, even when they contain foreign DNA.
- They contain a cloning site to insert foreign DNA (restriction enzyme sites in non-essential regions).
- They code for a phenotypic trait that can be used to detect their presence, it is also possible to distinguish parental from recombinant vectors.

Plasmids, Bacteriophages, Cosmids, Yeast artificial chromosomes, Bacterial artificial chromosomes and Expression vectors will now be discussed in more detail.

**Plasmids**

The first cloning vectors to be used, in the mid-1970s, were naturally occurring bacterial plasmids, originally from *Escherichia coli*. Plasmids are small, extra-chromosomal, circular DNA molecules that autonomously replicate inside the bacterial cell. They are convenient for the cloning of small DNA fragments (up to 20 kb). They contain an origin of replication (ori), which enables them to be replicated independently, although this normally relies on polymerases and other components of the host cell’s machinery (Figure 2). They usually carry a few genes, one of which may confer resistance to antimicrobial substances, which can be used as a selectable marker. The most widely known resistance gene is the *bla* or *amp*’ gene, encoding β-lactamase, which degrades penicillin antibiotics such as ampicillin. Another is the *tetA* gene, which encodes a transmembrane pump able to remove the antibiotic tetracycline from the cell (Turner *et al.*, 1997).

![Figure 2: Diagram of a plasmid indicating origin of replication (ori), restriction sites and gene for antibiotic resistance (selectable marker).](image)

**Bacteriophages**

Bacteriophages, or phages, are viruses that specifically infect bacteria. Like all viruses, phages are very simple in structure, consisting merely of a DNA (or occasionally RNA) molecule carrying a number of genes, including several for replication of the phage, surrounded by a protective coat or capsid made up of protein molecules (Figure 3). Both single-stranded (filamentous) and double-stranded phages have been exploited as cloning vectors (Chauthaiwale *et al.*, 1992). Bacteriophage
Lambda (phage λ), a head- and tail double stranded phage vector (Figure 3a), is probably the best-understood viral vector. Phage λ infects *E. coli* and undergoes either a lytic or lysogenic life cycle.

![Diagram of phage structure](image)

**Figure 3:** Schematic representation of the two main types of phage structure. (a) Head-and-tail (e.g. phage λ), (b) Filamentous (e.g. M13) (Brown, 1990).

The general pattern of infection:

- The phage particle attaches to the outside of the bacterium and injects its linear DNA molecule into the host cell.
- The phage DNA molecule is replicated, usually by specific phage enzymes coded by genes on the phage chromosome.
- Other phage genes direct synthesis of the protein components of the capsid, and new phage particles are assembled and released from the bacterium by lysis and cell death (the lytic phase).

In contrast to the lytic cycle, the lysogenic phase is characterized by retention of the phage DNA molecule in the host bacterium. The phage DNA will integrate into the host genome by site-specific recombination, where it may remain for a long period. The integrated form of the phage DNA (called the prophage) is quiescent, and a bacterium (referred to as a lysogen), which carries a prophage, is usually physiologically indistinguishable from an uninfected cell. However, the prophage is eventually released from the host genome and the phage reverts to the lytic mode and lyses the cell. The infection cycle of phage λ is summarized in Figure 4.
Lysogenic infection cycle: The pattern of phage infection that involves integration of the phage DNA into the host chromosome.

Lytic infection cycle: The pattern of infection displayed by a phage that replicates and lyses the host cell immediately after the initial infection. Integration of the phage DNA molecule into the bacterial chromosome does not occur.

The genes encoding the head and tail proteins of phage λ, as well as other genes involved in the lytic and lysogenic cycles, are clustered in distinct regions in the 50 kb phage λ genome. Genes involved in lysogeny and certain other genes irrelevant for lytic growth, can be deleted from the viral genome, and replaced by other DNA sequences of interest. This forms the basis of the use of phage λ as a cloning vector.

Cosmids:

Cosmids are plasmids containing phage lambda cos ends, they are 4 to 6 kb in size and are specifically designed for cloning of large DNA fragments (up to 45 kb). They have (i) a drug resistance marker (such as the ampicillin resistance gene), (ii) a plasmid origin of replication (ori), (iii) a fragment
carrying the ligated cohesive ends (cos) of phage λ, and (iv) one or more unique restriction sites for cloning (Collins & Hohn, 1978) (Figure 5). Recombinant cosmid molecules can be conveniently packaged in vitro inside a phage coat by the cleavage of two cos sites flanking the insert DNA. The resultant phages are then infected into a suitable E. coli host. (A cosmid molecule alone cannot be packaged because it falls short of the minimum size required for packaging). Inside a cell, two cos ends are ligated by the host ligase, resulting in a circular molecule which can be propagated as a plasmid, and a drug resistance marker is expressed. As cosmids lacks all the λ genes, plaques will not be produced, but instead colonies are formed on selective media (Turner et al., 1997; Brown, 1990)

![Figure 5: The basic features of a cosmid. OriV = origin of replication, Cos sites = provide blunt ends, EcoRI, SmaI = restriction endonuclease sites. (Adapted from: http://www.bio.davidson.edu/Courses/Molbio/MolStudents/spring2003/WilsonE/cosmid.gif).](http://www.bio.davidson.edu/Courses/Molbio/MolStudents/spring2003/WilsonE/cosmid.gif)

Yeast artificial chromosomes (YAC):

Yeast artificial chromosomes (YAC) were created in 1987 (Burke et al., 1987) to provide a general method of cloning megabase-sized DNA fragments. A YAC is a functional artificial chromosome (self-replicating element) and includes the following DNA sequences needed for replication in yeast cells:

- **TEL**: The telomere, which is located at each chromosome end, protects the linear DNA from degradation by nucleases,
- **CEN**: The centromere, which is the attachment site for mitotic spindle fibers, pulls one copy of each duplicated chromosome into each new daughter cell, and
- **ORI**: Replication origin sequences, which are specific DNA sequences to allow the DNA replication machinery to assemble on the DNA and move at the replication forks.

Additionally, it contains genes that code for selectable gene products (that allow the easy isolation of yeast cells that have taken up the artificial chromosome), and recognition sites for restriction enzymes (to be used as the cloning sites) (Burke et al., 1987).
Bacterial artificial chromosomes (BAC):

Bacterial artificial chromosomes (BAC) emerged in the early 1990s as an alternative to YACs (Shizuya et al., 1992). BACs are not real artificial chromosomes, but rather modified bacterial F factors (Peterson et al., 2000) and can carry up to 500 kb. They are based on the single-copy F-plasmid replicon containing four essential regions:

- An origin of replication \( (\text{oriS}) \),
- Loci involved in proper partitioning \( (\text{parA, parB and parC}) \),
- A chloramphenicol-antibiotic resistance marker, and
- \( \text{repE} \), encoding RepFIA protein E which is autoregulatory and essential for replication from \( \text{oriS} \). (Choi & Wing, 1998).

Expression vectors:

An expression vector is a cloning vector designed in such a manner that a foreign gene, inserted into the vector, will be translated and expressed into protein or protein fragments in the host organism. The type of host cell to be used depends on the protein to be expressed and includes bacteria, yeasts, cultured insect cells or cultured mammalian cells.

Expression systems are used for the artificial synthesis of proteins encoded by recombinant-DNA constructs and are useful for several reasons:

- The production of biotechnologically relevant proteins (enzymes, pharmaceuticals).
- The selection of clones via the specific detection of a gene product (e.g. by immunological screening).
- The study of the function of a gene product inside the cell that is used as an expression system.
- The analysis of interactions of the gene product with other molecules by testing expression products of gene fragments or by introducing mutations (e.g. epitope mapping).

Table 1 summarizes the basic features of the cloning vectors discussed.

<table>
<thead>
<tr>
<th>Vector</th>
<th>Size of insert</th>
<th>Appearance of recombinants</th>
<th>Common applications</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plasmid</td>
<td>20 kb</td>
<td>Colonies</td>
<td>Gene cloning</td>
</tr>
<tr>
<td>Phage Lambda</td>
<td>20 kb</td>
<td>Plaques</td>
<td>Gene cloning, cDNA libraries</td>
</tr>
<tr>
<td>Phage M13</td>
<td>5 kb</td>
<td>Plaques</td>
<td>Nucleotide sequencing, site-directed mutagenesis</td>
</tr>
<tr>
<td>Cosmids</td>
<td>45 kb</td>
<td>Colonies</td>
<td>Cloning of large DNA fragments</td>
</tr>
<tr>
<td>YACs</td>
<td>More than 1 Mbp</td>
<td>Yeast cells</td>
<td>Cloning of megabase-sized DNA fragments, Gene libraries</td>
</tr>
<tr>
<td>BACs</td>
<td>500 kb</td>
<td>Colonies</td>
<td>Gene libraries, sequencing of complete genomes</td>
</tr>
<tr>
<td>Expression vectors</td>
<td>Depends on vector</td>
<td>Colonies or plaques</td>
<td>Gene expression, expression libraries</td>
</tr>
</tbody>
</table>
RESTRICTION ENDONUCLEASES

To incorporate fragments of foreign DNA into a cloning vector, methods for cutting and rejoining of single stranded DNA are necessary. The identification of restriction endonucleases in the 1960s and early 1970s and the recognition that these enzymes act as "molecular scissors", always cutting DNA at specific locations (base sequences), was the key discovery which allowed the cloning of DNA to become a reality (Turner et al., 1997). Restriction enzymes are proteins that bind to a specific DNA sequence and cleave (hydrolyze) the backbone of the DNA between the deoxyribose and phosphate groups (the phosphodiester linkage). This results in phosphate groups on the 5′ ends and hydroxyl groups on the 3′ ends of both strands. The biological function of restriction enzymes is to protect the bacterial cell against the introduction of foreign DNA into the cell (Turner et al., 1997). DNA methylases also exist in the cell to prevent the restriction enzymes from digesting the cell’s own DNA. This is done by the modification of the DNA at specific sites which prevents the restriction enzymes from binding or cleaving the DNA.

Restriction enzymes are distinguished according to their recognition sequence and their cleavage site. Three main groups have been recognized:

- **Type I** enzymes that cut DNA up to 5 kbp away from the recognition site,
- **Type II** enzymes which cut DNA inside the recognition site, and
- **Type III** that cut DNA just a few base pairs away from the recognition site (also called shifters).

In recombinant DNA technology, the lengths of the fragments generated, and the composition of the termini of the fragments generated are the two important criteria to consider in the choice of a restriction enzyme. The length of the restriction fragments generated will depend on several factors, including (i) the length of the recognition sequence and its composition, (ii) the G+C content of the organism’s DNA, and (iii) its methylation status (most restriction enzymes do not recognize methylated palindromes and do not cleave these sequences). With regards to termini generated upon restriction, three types are differentiated:

- **5′-extensions** (5′-sticky or cohesive ends) as generated by most restriction enzymes, eg. EcoRI (G’AATTC)
- **3′-extensions** (3′-sticky or cohesive ends) as generated by others such as PstI (CTGCA’G)
- **Blunt or flush ends** produced for example by PvuII (CAG’CTG) or HaeII (GG’CC).

These features are crucial and have to be taken into consideration when manipulating DNA since nucleases and polymerases recognize certain types of ends only. When joining (ligating) together restriction fragments, it is important to have complementary sticky ends, which facilitate the annealing process and, therefore, increase ligation efficiency (http://www.brunel.ac.uk/depts/bl/project/genome/moltec/).
Restriction enzymes as “molecular scissors”:
(http://www.hort.purdue.edu/hort/courses/HORT250/lecture%2003)

1. Restriction enzymes recognize a specific sequence of bases in a DNA molecule. In the example shown below, the enzyme EcoRI finds the sequence G’AATTC.

2. The restriction enzyme then cuts through the two strands of DNA molecule in a very precise manner, indicated by the red line. Fragments of DNA produced by restriction enzymes can now be used for cloning. Note that this enzyme has cut the DNA to leave protruding sticky ends.

3. Some other restriction enzymes cut DNA to produce blunt ends, for example HaeIII:

A restriction digest will result in a number of DNA fragments, the sizes of which depend on the exact positions of the recognition sequences for the endonuclease in the original DNA molecule. These DNA fragments often require separation and/or analysis. Gel electrophoresis is frequently used to separate fragments of DNA (or RNA) according to their size and is commonly used to identify specific cloned genes.

LIGATION: THE JOINING OF DNA MOLECULES

The final step in the construction of a recombinant DNA molecule is the joining together of the vector molecule and the foreign (target) DNA to be cloned. This process is referred to as ligation, and the enzyme that catalyses the reaction is called DNA ligase (Brown, 1990).

All living cells produce DNA ligases and within the cell these enzymes carry out the important function of repairing any discontinuities in double-stranded DNA molecules. Such a discontinuity is simply a position where a phosphodiester bond between adjacent nucleotides is missing. Although discontinuities may arise by chance breakage of the cell’s DNA molecules, they are also a natural result of processes such as DNA replication and recombination. Ligases, therefore, play several vital roles in the cell. In a test-tube, purified DNA ligases (eg. T4 DNA ligase, derived from the T4 bacteriophage) will join together individual DNA molecules or the two ends of the same molecule by
catalyzing the formation of a phosphodiester bond between the 5’ phosphate of one strand and the 3’ hydroxyl of the other (Brown, 1990).

There are three main methods of joining DNA fragments (www.users.wmin.ac.uk/~redwayk/lectures/+):

- **Using sticky (cohesive) ends.** This is the easiest and most commonly used method. Fragment ends must be complementary. Therefore, the same restriction enzyme is used to cut both vector and insert DNA, or two different enzymes are used that produce complementary ends.
- **Using synthetic linkers (adaptors).** Ligases can join blunt ends, but at a lower efficiency. Linkers are chemically synthesized pieces of double-stranded DNA that include a single restriction site for an enzyme that produces sticky ends. DNA ligase will attach linkers to blunt-ended DNA molecules, thus enabling the conversion of blunt ends to sticky ends (after restriction digestion).
- **Using homopolymer tailing.** This technique offers a different approach to the production of sticky ends on a blunt-ended DNA molecule. Long tails of complementary-pairing bases (G and C or T and A) are added to both vector and insert DNA respectively by using the terminal transferase enzyme.

## INTRODUCTION OF DNA INTO HOST CELLS

The next step in a gene cloning experiment is to introduce the recombinant molecule into living cells (usually bacteria), which will then grow and divide to produce clones (Brown, 1990).

**Transformation:** The uptake of DNA by bacterial cells
Transformation is the process of uptake of foreign DNA (normally plasmids) by bacteria. In nature, transformation is probably not a major process by which bacteria obtain genetic material. This is reflected by the fact that only a few species (notably members of the genera *Bacillus* and *Streptococcus*) can be transformed with ease in the laboratory, and studies have revealed that these species possess sophisticated mechanisms for DNA binding and uptake. Under normal circumstances, most bacterial species (including *E. coli*) will only take up limited amounts of DNA. In order to transform these species efficiently, the bacteria have to undergo some form of physical and/or chemical treatment that will enhance their ability to take up DNA. Cells that have undergone treatment are referred to as competent cells (Brown, 1990; Turner *et al.*, 1997).

It was discovered in the early 1970s (cited by Old & Primrose, 1994) that *E. coli* cells treated with solutions containing Ca$^{2+}$ ions were rendered susceptible to take up exogenous DNA. The precise mechanism of this is not understood. Ca$^{2+}$ ions possibly cause the DNA to precipitate on the outside of the cells, or perhaps they are responsible for some kind of change in the cell wall that improves DNA binding. Thus, treatment with Ca$^{2+}$ ions affects only DNA binding, and not the actual uptake of DNA (Brown, 1990). Furthermore, when DNA is added to treated cells, it remains attached to the exterior, and is not at this stage transported into the cytoplasm. The actual movement of the DNA into competent cells is stimulated by briefly rising the temperature to 42°C (referred to as heat shock) (Figure 6). This process induces enzymes involved in the repair of DNA and other cellular components, which allow the cell to recover from the unusual conditions of the transformation process, and increases the efficiency. After this, cells are incubated in a growth medium and finally spread out on an agar plate and incubated until single colonies of bacteria grow (Brown, 1990; Turner *et al.*, 1997).

![Figure 6: Binding and uptake of foreign DNA by a competent cell.](image-url)
The quality of a given preparation of competent cells may be measured by determining the transformation efficiency. This can be defined as the number of colonies formed (on a selective plate) per microgram of input DNA, where that DNA is a pure plasmid, most commonly the vector to be used in a cloning experiment. Transformation efficiencies can range from $10^3$ per μg for crude transformation protocols, to more than $10^8$ per μg for very carefully prepared competent cells to be used in the construction of libraries. A transformation efficiency of $10^5$ per μg would be adequate for a simple cloning experiment (Turner et al., 1997).

**Introduction of phage DNA into bacterial cells:**

A recombinant DNA molecule, constructed with a phage vector, can be introduced into a bacterial cell by transfection or *in vitro* packaging.

**Transfection**

The process of transfection is equivalent to transformation, the only difference being that phage DNA rather than a plasmid is involved. As with transformation, the purified recombinant phage molecule is added to competent *E. coli* cells and DNA uptake is induced by heat shock (Brown, 1990).

**In vitro packaging**

Another method of introducing a recombinant DNA molecule, constructed with a phage vector, is *in vitro* packaging. To introduce such recombinants into the host cell, they are first “packaged” into phage particles *in vitro*, and then the bacteria are infected (transfected) with the packaged recombinant (Adams et al., 1992; Brown, 1990).

**Transformation of non-bacterial cells:**

Ways of introducing DNA into yeast, fungi, animals and plants are also needed if these organisms are to be used as hosts for gene cloning (Brown, 1990). The uptake of DNA into eukaryotic cells is more challenging than bacterial transformation, and the efficiency of the process is much lower. For example, in yeast and plant cells the cell wall must be digested with degradative enzymes to yield fragile protoplasts, which may then take up DNA quite readily. The cell walls are re-synthesized once the degrading enzymes are removed. In contrast, animal cells in culture, which have no cell wall, will take up DNA at low efficiency if it is precipitated on their surface with calcium phosphate. Treating the cells with a high voltage, which is believed to open transient pores in the cell membrane, may increase the efficiency of the process. This process is referred to as electroporation and can also be used to introduce cloned genes into bacterial cells. In addition, direct physical methods have been used. DNA may be microinjected directly into the cytoplasm or even the nucleus of individual animal or plant cells in culture. Alternatively, DNA may be introduced by the bombardment of the target cells with metallic microprojectiles coated with DNA. This technique is referred to as biolistics (Brown, 1990; Turner et al., 1997).
GENE LIBRARIES

A gene library is a collection of different DNA sequences from an organism each of which has been cloned into a vector for ease of purification, storage and analysis. Gene libraries are used for two main purposes: (i) the cloning of individual genes (e.g. genes associated with certain genetic disorders), and (ii) the construction of physical maps of genomes (Brown, 1990). There are essentially two types of gene libraries that can be made depending on the source of DNA used. If the DNA is genomic DNA, the library is called a **genomic library**. If the DNA is a copy of an mRNA population, that is cDNA, then the library is called a **cDNA library** (Brown, 1990; Turner et al., 1997).

**Basic steps in the construction of a gene library:**

1. Extraction of genomic DNA of the organism of interest.
2. The extracted DNA is cut into gene-size pieces with restriction enzymes.
3. The appropriate vector (for this purpose a plasmid) is cut with the same restriction enzyme.
4. The gene-sized DNA and cut plasmids are combined into one test tube. Some of the enzyme cut DNA pieces will combine with the cut vectors and form recombinant DNA.
5. The recombinant plasmids are then transferred into bacteria using either electroporation or heat shock (transformation).
6. The bacteria are grown on a culture dish and allowed to grow into colonies. All the colonies on all the plates (cultures) are called a gene library.
7. The gene library is then screened in order to discover which bacterial colony is making copies of the one gene you are interested in. Library screening identifies colonies which have that particular gene. Screening can be based on detecting the DNA sequence of the cloned gene, detecting a protein that the gene encodes, or the use of linked DNA markers.
8. When the bacterial colony containing the desired gene is located, the bacteria can be propagated to make millions of copies of the recombinant plasmid that contains the gene. Furthermore, the plasmids can be extracted for the next steps of genetic engineering, gene modification, and transformation.

SCREENING AND SELECTION FOR RECOMBINANTS

The art of cloning is to find the one particular transformed cell that contains the cloning vector with the gene of interest (referred to as a recombinant cell). It is thus most important to be able to select recombinant cells from transformed cells (containing the vector without insert DNA) (Brown, 1990). If this goal has been achieved the gene in question is said to have been cloned.

Although there are many different procedures by which the desired clone can be obtained, all are variations of two basic concepts (Brown, 1990):

- **Direct selection**, which means that the cloning experiment is designed in such a way that the clones obtained are the clones containing the required gene. Almost invariably, selection occurs
at the plating-out stage. Direct selection is the method of choice, as it is quick and usually unambiguous. However, it is not applicable to all genes.

- **Identification of the clone from a gene library**, which entails an initial “shotgun” cloning experiment, to produce a clone library representing all or most of the genes present in the cell, followed by analysis of the individual clones to identify the correct one.

**Direct selection:**

To be able to select for a cloned gene, it is necessary to plate the transformants on to an agar medium on which only the desired recombinants, and no others, can grow. The only colonies that will appear will, therefore, comprise cells that contain the desired recombinant DNA molecule. Most cloning vectors are designed so that the insertion of a DNA fragment into the vector destroys the integrity of one of the genes present on the molecule. This is referred to as insertional inactivation (Brown, 1990). Two examples will be discussed: (i) direct antibiotic resistance screening, and (ii) blue-white colour screening.

**Direct antibiotic resistance screening**

Most cloning vectors are designed so that the insertion of a DNA fragment into the vector destroys the integrity of one of the genes present on the molecule (usually an antibiotic resistance gene). For instance, if the vector carries an ampicillin resistance gene ($\text{amp}^\text{R}$), it will confer ampicillin resistance to the $E. \text{coli}$ cells if the transformation process was successful. This would mean that when the clones are plated out on ampicillin-containing medium, only cells containing a plasmid would be resistant to the antibiotic and, therefore, be able to grow. However, this does not show whether the cell contains the recombinant plasmid (with the inserted gene of interest) or only the re-ligated vector (without the inserted gene of interest). Figure 7 shows a diagram of screening for recombinants by using direct antibiotic resistance.
Blue-white colour screening

A more sophisticated procedure for screening for the presence of recombinant plasmids, which can be carried out on a single transformation plate, is called blue-white screening. This method also involves the insertional inactivation of a gene and uses the production of a blue compound as an indicator (Brown, 1990). The gene is lacZ, which encodes the enzyme β-galactosidase, and is under the control of the lac promoter. If the host E. coli strain is expressing the lac repressor, expression of a lacZ gene on the vector may be induced using IPTG (isopropyl- β-D-thiogalactopyranoside), and the expressed enzyme can utilize the synthetic substrate X-gal (5-bromo-4-chlore-3-indolyl- β-D-galactopyranoside) to yield a blue product. Insertion of a DNA fragment into the lacZ gene (insertional inactivation of lacZ) in the production of a recombinant plasmid would prevent the development of the blue colour. In this method, the transformed cells are spread onto a plate containing an antibiotic (to select for transformants in the usual way), IPTG and X-gal, to yield a mixture of blue and white colonies. The white colonies have no expressed β-galactosidase and are hence likely to contain the inserted target fragment. The blue colonies probably contain religated vector.

Identification of the clone from a gene library:

Many different techniques are available for screening a library. The most important approaches involve screening by nucleic acid hybridization and screening by functional analysis. Nucleic acid hybridization requires some prior knowledge of the DNA sequence either of the gene to be cloned or
of stretches of DNA in the vicinity of the gene to be cloned. Functional screening involves the use of expression vectors that allow cells containing the vector with the desired gene to express the corresponding protein. Under these circumstances, cells containing a vector with the desired gene can be identified by means of antibodies directed against the protein. Cells producing the desired gene product can also be identified in bioassays detecting protein activities, if these are available (http://www.copewithcytokines.de/).

**Nucleic acid hybridization**

Detection of an individual clone in a library can be achieved by employing strategies of nucleic acid hybridization in which short chemically synthesized labeled oligonucleotides (probes) are used to detect complementary sequences in individual cells or phages containing an insert. The success of colony or plaque hybridization will depend on the availability of a DNA molecule that can be used as probe. This probe must share at least a part of the sequence of the cloned gene. If the gene itself is not available, it can be derived, for example, from known protein sequences, or so-called degenerate oligonucleotides (mixtures of oligonucleotides that differ from each other by base substitutions at identical and/or different positions) can be used.

The first step of a hybridization screening experiment involves the transfer of the DNA in the plaque or colony to a nylon or nitrocellulose membrane. The bacteria on the membrane are lysed to release their DNA, and the DNA is denatured with alkali to produce single strands that are bounded to the membrane by heat treatment or UV irradiation. The membrane is then immersed in a solution containing a nucleic acid probe (usually radioactively labeled) and incubated to allow the probe to hybridize to its complementary sequence. After hybridization, the membrane is washed to remove unhybridized probe, and regions where the probe has hybridized are visualized with autoradiography (Turner et al., 1997). Figure 8 shows a schematic sketch of a hybridization screening experiment.

![Figure 8: An example of hybridization probing (Hybridization screening of a plasmid cDNA library using a radioactive probe) (Adapted from: http://www.genetics.biol.ttu.edu/genetics/Lecture/red28.html)](http://www.genetics.biol.ttu.edu/genetics/ Lecture/red28.html)
Functional screening

The desired gene can also be identified by the activities of the encoded gene product. In this case, one uses a so-called expression library that has been established by cloning DNA (cDNA in the case of eukaryotes) fragments into special cloning vectors allowing the functional expression of cloned DNA fragments. Functional gene products and hence the desired clones can then be detected either by antibodies (immunological screening) or other ligands that specifically recognize the encoded proteins or by exploiting a bioactivity of the gene product, if known (Brown, 1990).

The procedure of immunological screening has similarities to hybridization screening, discussed before, though in this instance it is the protein encoded by the DNA (or cDNA) rather than the DNA itself that is detected on the membrane. The membrane is treated to covalently attach the protein, and immersed in a solution containing the antibodies. When the antibody has bound to its epitope, it is detected by other antibodies and/or chemicals that recognize it.

Chromosome walking

Often, a genomic clone may not include all of the sequence for a particular gene so it is necessary to isolate overlapping clones that cover the genomic region of interest. Such clones can be isolated using probes derived from the outlying ends of the clones in hand. This process is known as chromosome walking.

ANALYSIS OF A CLONE

Once a clone containing a target gene is identified, and before it is used in any of the diverse applications of DNA cloning, various properties of the recombinant DNA molecule, such as size, restriction map, orientation of the gene present and the nucleotide sequence thereof, can be determined. Before sequencing became automated (and so rapid), it was common to perform all these steps of characterization, but nowadays sequencing all or part of the clone is often the first step. However, while the sequence will provide the size, restriction map and the orientation of the predicted gene(s), or open reading frame (ORF), experiments must be performed to verify that the predicted gene is actually transcribed into RNA in some cells of the organism and that, for protein coding genes, a protein of the predicted size is made \textit{in vivo} (Turner \textit{et al.}, 1997).

APPLICATIONS

Gene, or molecular, cloning, the process of creating genetically identical DNA molecules, provides the foundation of the molecular biology revolution and is a fundamental and essential tool of biotechnology research, development and commercialization. Virtually all applications of recombinant DNA technology, from the Human Genome Project, to pharmaceutical manufacturing, to the
production of transgenic crops, depend on molecular cloning. Currently, recombinant DNA technology, in conjunction with molecular cloning, plays an important role in:

- Health care (production of new medicines or therapeutics, production of safer vaccines, diagnostics, and gene therapy)
- Agriculture (improvement of agricultural crops, cultivation of transgenic crops with traits such as herbicide tolerance, insect resistance and disease resistance)
- The process industry (improving existing proteins, such as enzymes, antibodies and cell receptors, creating proteins not found in nature - the chemical, textile, pharmaceutical, pulp and paper, food and feed, and energy industries are all benefiting from such proteins)
- Environmental management/biotechnology (the use of naturally occurring microorganisms to identify and filter manufacturing waste before its introduction into the environment, genetically modified microorganisms in waste treatment and pollution prevention, etc)

**DEFINITIONS**

**Bacterial artificial chromosome (BAC):** BACs are a type of cloning vector. They are not real artificial chromosomes, but rather modified bacterial F factors and can carry up to 500 kb.

**Bacteriophage:** A virus for which the natural host is a bacterial cell. Used as a vector for cloning segments of DNA.

**Biolistics:** A means of introducing DNA into cells that involves bombardment with high-velocity micro-projectiles coated with DNA.

**Blunt end:** An end of a DNA molecule at which both strands terminate at the same nucleotide position with no single-stranded extension.

**Capsid:** The protein coat that encloses the DNA or RNA molecule of a bacteriophage or virus.

**Chimera:** A recombinant DNA molecule made up of DNA fragments from more than one organism (named after the mythological beast).

**Chromosome:** A self-replicating nucleic acid molecule carrying a number of genes.

**Chromosome walking:** The repeated screening of a genomic library to obtain overlapping clones and hence build up a collection of clones covering part of a chromosome. It involves using the end of the inserts as probes for subsequent rounds of screening. Also, a technique used to identify a series of overlapping restriction fragments, often to determine the relative positions of genes on large DNA molecules.

**Clone:** A gene or DNA fragment, and/or the collection of cells or organisms (such as bacteria) containing the cloned piece of DNA. A large number of cells or molecules that is identical to a single parental cell or molecule.
Molecular Biology: Gene cloning

**Competent cells:** Refers to a culture of bacteria that have been treated to enhance their ability to take up DNA molecules.

**Complementary DNA (cDNA) cloning:** A cloning technique involving conversion of purified mRNA to DNA before insertion into a vector.

**Cos site:** One of the cohesive, single-stranded extensions present at the ends of the DNA molecules of certain strains of phage lambda.

**Cosmid:** A cloning vector consisting of the phage lambda cos sites inserted into a plasmid. Used to clone fragments up to 45 kb in size.

**DNA libraries:** DNA libraries, consisting of sets of random cloned fragments of either genomic or cDNA, each in a separate vector, are used in the isolation of unknown genes.

**DNA sequencing:** Determination of the order of nucleotides in a DNA molecule.

**Electrophoresis:** Separation of molecules based on their net electric charge.

**Electroporation:** A method for increasing DNA uptake by bacteria or protoplasts through prior exposure to a high voltage that results in the temporary formation of small pores in the cell membrane.

**Endonuclease:** An enzyme that breaks phosphodiester bonds within a nucleic acid molecule.

**Exonuclease:** An enzyme that sequentially removes nucleotides from the ends of a nucleic acid molecule.

**Expression vector:** A cloning vector designed so that a foreign gene inserted into the vector will be expressed in the host organism.

**Gene:** A segment of DNA that codes for an RNA and/or polypeptide molecule.

**Gene cloning:** Insertion of a fragment of DNA, carrying a gene, into a cloning vector, and subsequent propagation of the recombinant DNA molecule in a host organism.

**Gene manipulation:** The formation of new combinations of heritable material by the insertion of nucleic acid molecules, produced by whatever means outside the cell, into any virus, bacterial plasmid or other vector system so as to allow their incorporation into a host organism in which they do not naturally occur but in which they are capable of continued propagation.

**Gene mapping:** Determination of the relative positions of different genes on a DNA molecule.

**Genome:** The complete set of genes of an organism.

**Genomic library:** A collection of clones sufficient in number to include all the genes of a particular organism.
Homopolymer tailing: The attachment of a sequence of identical nucleotides (e.g. AAAAA) to the end of a nucleic acid molecule, usually referring to the synthesis of single-stranded homopolymer extensions on the ends of a double-stranded DNA molecule.

Hybridization probe: A labeled nucleic acid molecule that can be used to identify complementary or homologous molecules through the formation of stable base-paired hybrids.

Immunological screening: The use of an antibody to detect a polypeptide synthesized by a cloned gene.

Insertional activation: The cloning strategy whereby insertion of a new piece of DNA into a vector inactivates a gene carried by the vector.

In vitro packaging: Synthesis of infective lambda particles from a preparation of lambda capsid proteins and a concatamer of DNA molecules separated by cos sites.

Lambda (λ): A bacteriophage that infects E. coli, derivatives of which are extensively used as cloning vectors.

Ligase (DNA ligase): An enzyme that repairs single-stranded discontinuities in double-stranded DNA molecules in the cell. Purified DNA ligase is used in gene cloning to join DNA molecules together.

Linker: A synthetic, double-stranded oligonucleotide used to attach sticky ends to a blunt-ended molecule.

Lysogen: A bacterium that harbours a prophage.

Lysogenic infection cycle: The pattern of phage infection that involves integration of the phage DNA into the host chromosome.

Lytic infection cycle: The pattern of infection displayed by a phage that replicates and lyses the host cell immediately after the initial infection. Integration of the phage DNA molecule into the bacterial chromosome does not occur.

M13: A filamentous bacteriophage that infects E. coli, derivatives of which are extensively used as cloning vectors.

Nucleic acid hybridization: Formation of a double-stranded molecule by base pairing between complementary or homologous polynucleotides.

Origin of replication (ori): The specific position on a DNA molecule where DNA replication begins.

Plaque: A zone of clearing on a lawn of bacteria caused by lysis of the cells by infecting phage particles.

Plasmid: A small, extra-chromosomal, circular molecule of DNA that replicates independently of the host DNA.

Prophage: The integrated form of the DNA molecule of a lysogenic phage.
Protoplast: A cell from which the cell wall has been completely removed.

Recombinant: A transformed cell that contains a recombinant DNA molecule.

Recombinant DNA: A DNA molecule produced by inserting DNA from one organism (or may be chemically synthesized) into another piece of DNA using gene manipulation techniques.

Recombinant protein: A polypeptide that is synthesized in a recombinant cell as the result of expression of a cloned gene.

Recombination: The exchange of DNA sequences between different molecules, occurring either naturally or because of DNA manipulation.

Replica plating: A technique whereby the colonies on an agar plate are transferred to a new plate, on which the colonies will grow in the same relative positions as before.

Restriction analysis: Determination of the number and sizes of the DNA fragments produced when a particular DNA molecule is cut with a particular restriction endonuclease.

Restriction endonucleases: Bacterial enzymes that cut (hydrolyze) DNA at specific recognition sequences into defined and reproducible fragments. In bacteria, they form part of the restriction-modification defence mechanism against foreign DNA.

Selection: A means of obtaining a clone containing a desired recombinant DNA molecule.

Sticky-ended DNA: DNA with complementary single-strands that protrude from opposite ends of a DNA duplex.

Transduction: The process of introducing genetic information into a cell by encapsulating the nucleic acid in a viral or phage coat.

Transfection: The introduction of purified phage DNA molecules into a bacterial cell.

Transformation: Transformation is the process of take-up of foreign DNA, normally plasmids, by bacteria.

Transformation frequency / efficiency: A measure of the proportion of cells in a population that are transformed in a single experiment.

Vector: An agent or vehicle that carries a DNA fragment into a host cell. If it is used for reproducing the DNA fragment, it is called a cloning vector. If it is used for expressing a certain gene in the DNA fragment, it is called an expression vector.

Yeast Artificial Chromosome (YAC): A cloning vector comprising the structural components of a yeast chromosome and able to clone very large pieces of DNA.
REFERENCES