Ministry of Higher Education and Scientific Research
University of Baghdad
College of Science
Department of Biology

Genetic engineering
2021-2022

المرحلة الرابعة - الدراستين الصباحية والمسائية

الفصل الدراسي الثاني

تدريسي المادة :

أ.م.د. لبنى محي رسول

م.د. محمد عبد الرحمن محمد

Lecture 1

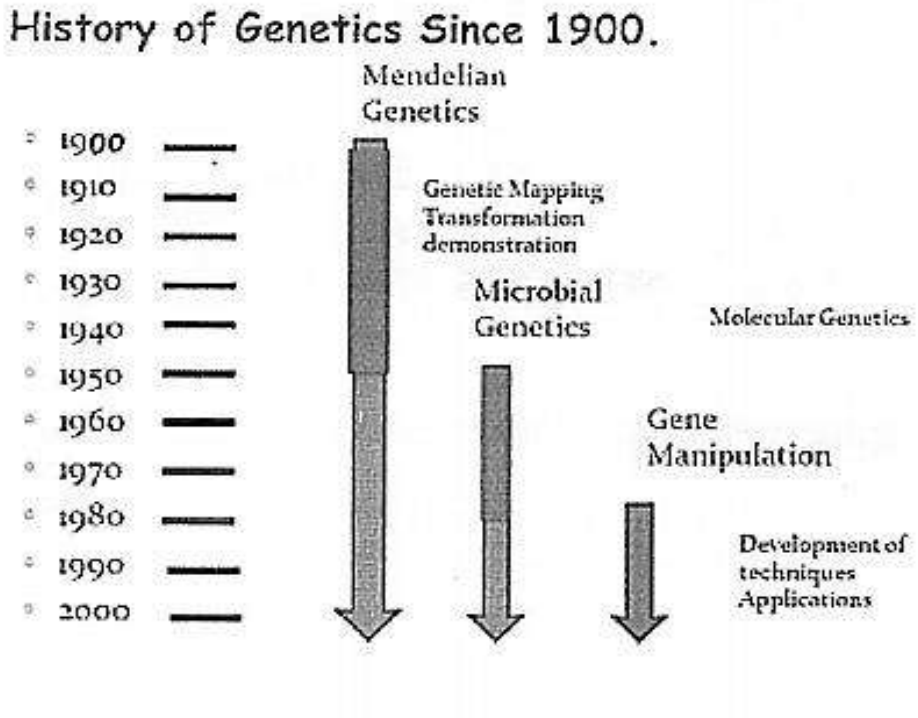
Introduction

Genetic engineering involves the group of techniques used to cut up and join together genetic material, especially DNA from different biological species, and to introduce the resulting hybrid DNA into an organism in order to form new combinations of heritable genetic material.

However, there are several other terms that can be used to describe this technology, including gene manipulation, gene cloning, recombinant DNA technology, genetic modification, and the new genetics.

Development of this technology began in 1972. An organism that is generated through genetic engineering is considered to be a genetically modified organism (GMO).

Event	Year
first recombinant DNA molecules	1972
manipulation of genetic material in vitro	1970s
The first GMO were bacteria (inserting antibiotic resistance genes into the plasmid of an E. coli)	1973, 1974
a transgenic mouse was produced making it the world's first transgenic animal.	1978
Human insulin was the first production of genetically engineered in E. coli. The first field trials of genetically engineered plants occurred in France and the USA, Tobacco plants were engineered to be resistant to herbicides.	1986
11 transgenic crops were grown commercially in 25 countries.	2009
The first synthetic genome was created and inserted it into an empty bacterial cell.The resulting bacterium, named Synthia, could replicate and produce proteins.	2010



Basic concepts of Genetic Engineering: The technology is a set of methods used to locate, analyze, alter, study, transfer, and recombine DNA sequences. It is used to investigate the structure and function of genes, to answer the questions in many biological fields, produce commercial products, and to diagnose and treat diseases.

Application of genetic engineering

Many areas in which genetic engineering is of value, including the following:

1. Applications in molecular biology:

a- Basic research on gene structure and function.

b- Genome analysis by DNA sequencing such as complete human genome sequence project in 2003

2. Applications in medicine:

a- Production of recombinant pharmaceuticals: genetic engineering has produced commercially valuable proteins and gene therapies such as human insulin, human growth hormone and vaccines (Hepatitis B virus vaccine and edible vaccine).

b- Diagnosis of disease such as used of DNA probe and PCR technique to detect of viruses, bacteria and protozoa.

3.**Applications in agriculture:** plants have been modified for insect protection, herbicide resistance, virus resistance, enhanced nutrition, increasing yields tolerance to environmental pressures and the production of edible vaccines.

4.**Applications in zoology:** Genetically modified animals have been used for research. Producing animals with desirable traits such as animals with increased susceptibility to disease, hormones for extra growth and the ability to express proteins in their milk.

Although genetic engineering has the potential benefit in medicinal and agricultural fields, most people are concerned with possible risks of genetic engineering especially those associated with transgenic species.

Nucleases

A restriction nuclease is an enzyme that cuts the DNA molecule at, or near to, a specific nucleotide sequence to produce discrete DNA fragments that can be separated by gel electrophoresis.

This enzyme is capable of cleaving the phosphor-diester bonds between the nucleotide subunits of nucleic acids. Based on their mode of action, two main classes of nucleases have been defined: exonucleases are active at the end of nucleic acid molecules, and endonucleases cleave nucleic acids internally. Well known nucleases are deoxyribonucleases and Ribonuclease

Nucleases are double edge swords for molecular biologists: on one hand, they are the worst enemy to nucleic acids integrity and, on the other hand, they are very useful to cut and manipulate nucleic acids for cloning purposes. Many types of nucleases have been identified but the most widely used are DNase I and RNase A.

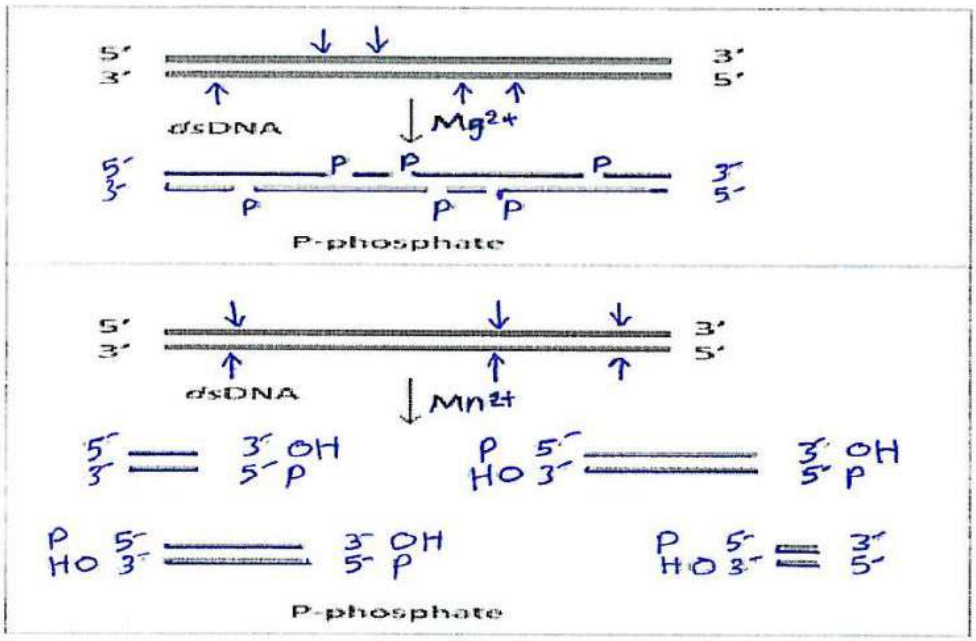
Deoxyribonucleases

Deoxyribonucleic (DNase is any enzyme that capable of hydrolyzing phosphor-diester bonds in the DNA backbone. A wide variety of deoxyribonucleases are known, which differ in their substrate specificities, chemical mechanisms, and biological functions. Some are indiscriminate the DNA sequence at which they cut, while others, including restriction enzymes, are very sequence-specific. Some cleave only double-stranded DNA; others are specific for single-stranded molecules; and still others are active toward both.

Deoxyribonuclease I (DNase I) is an endonuclease that acts on single or double stranded DNA. DNase I catalyzed cleavage occurs preferentially in 3' of a pyrimidine (C or T) nucleotide, and generates polynucleotides with free 3'-OH group and a 5'-phosphate.The nature of the divalent

cations present in the incubation mixture affects both specificity and mode of action of DNase I. The maximal activity of enzyme is obtained in the presence of Mg2+, and Mn2+ ions. In the presence of Mg2+, DNase I hydrolyzes each strand of duplex DNA independently, generating random cleavages while in the presence of Mn2+, DNase I cleaves both DNA strands at approximately the same site, producing blunt ends or fragments with 1-2 base overhangs. Some applications of DNase 1 are:-

- A recombinant form of this enzyme is used to treat one of the symptoms of cystic fibrosis by hydrolyzing the extracellular DNA in sputum and reducing its viscosity to clear the lung.
-DNasel is the most widely used enzyme in cloning experiments to remove DNA contamination from mRNA
-DNAse I is used for nick translation of DNA, for generating random preparation. augments for dideoxy sequencing.



Ribonuclease

Ribonuclease (RNase) is a type of nuclease that catalyzes the degradation of RNA into smaller components. RNases play key roles in the maturation of all RNA molecules, as well as cleaning of cellular RNA that is no longer required. In addition, active RNA degradation systems are the first defense against RNA viruses.

Pancreatic Ribonuclease (RNase A) is an endonuclease that cleaves single stranded RNA at 3' ending in C or U. RNase A is available from many commercial sources.

RNase A that commonly used in research is provided from bovine pancreatic. It is specific for single-stranded RNAs and it does not require any cofactors for its activity. RNase A is used to reduce RNA contamination in plasmid DNA preparations and for mapping mutations in DNA or RNA by mismatch cleavage.

Lecture 2

Restriction enzymes

Restriction enzymes, also known as restriction endonucleases, are enzymes that cut a DNA molecule at a particular site known as restriction site usually consist of four to eight nucleotides. In 1970, the first restriction enzyme was isolated from the bacterium *Haemophilus influenzae* strain Rd and showed that it was able to cleave DNA at specific sites. In 1972, another restriction enzyme, EcoRI, was isolated and characterized from Escherichia coli strain RY13. Now, more than 4,000 restriction enzymes are known (recognizing more than 300 distinct sequences) and more than 600 of these are available commercially. These enzymes are routinely used for DNA modification in laboratories, and represent one of the most important groups of enzymes for recombinant DNA technology.

Nomenclature

Restriction enzymes are named according to the bacterium from which they are purified. Briefly, three letters in italics are derived from the first letter of the genus and the first two letters of the microbial species from which the enzyme was derived. An additional letter without italics may be used to designate a particular strain. This is followed by a roman numeral to signify the first, second, etc., the enzyme discovered from the organism

Derivation of the *Eco*RI name		
Abbreviation	**Meaning**	**Description**
E	*Escherichia*	Genus
co	*coli*	Species
R	RY13	Strain
I	First identified	order of identification in the bacterium

Restriction-Modification system

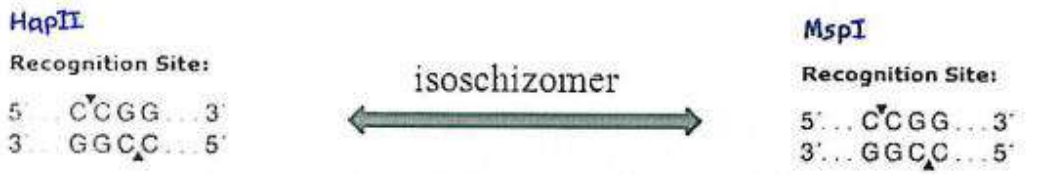
The restriction modification system (RM system) is found in bacteria and other prokaryotic organisms, and provides a defense against foreign DNA, such as that borne by bacterio phages.Restriction endonucleases were originally named for their ability to restrict the growth of bacteriophage in a host bacterial cell by cleavage of the invading DNA through restriction-modification system (R-M) that have two components include a restriction endonuclease and a

DNA methylase. These enzymes are found in bacterial cells, where they function as part of a protective mechanism.

Bacteria have restriction enzymes, also called restriction endonucleases, which cleave double stranded DNA at specific points into fragments, which are then degraded further by other endonucleases. This prevents infection by effectively destroying the foreign DNA introduced by an infectious agent (such as abacteriophage). Approximately one-quarter of known bacteria possess RM systems and of those about one-half have more than one type of system.

As the sequences recognized by the restriction enzymes are very short, the bacterium itself will almost certainly contain some within its genome. In order to prevent destruction of its own DNA by the restriction enzymes, methyl groups are added. These modifications must not interfere with the DNA base-pairing, and therefore, usually only a few specific bases are modified on each strand.

In this system the restriction enzyme hydrolyses any exogenous DNA that appears in the cell to protect bacteria from hydrolyze exogenous DNA that is not methylated by the host modification enzyme. Once methylated, the host DNA is no longer a substrate for the endonuclease, because both strands of the host DNA are methylated and even hemi-methylated DNA is protected. For example, *Haemophilus influenzae*, like all organisms that produce restriction enzymes, also produces enzymes that modify its own DNA so that it will not be cleaved by the restriction enzyme and E. coli RY13 strains, which produce EcoRI, also make EcoRI methylase. Isoschizomers and Neoschizomers. Isoschizomers are many enzymes from different biological sources recognize the same DNA sequence.



Neoschizomers are two enzymes recognize the same DNA sequence but cleave at a different position. for example *Sma*I (CCC/GGG) and *Xma*I (C/CCGGG)

Restriction site, or restriction recognition sites, are locations on a DNA molecule containing specific sequences of nucleotides(4-8 base pairs in length), which are recognized by restriction enzymes. These are generally palindromic sequences (because restriction enzymes usually bind as homodimers), and produce a double-stranded cut in the DNA. The recognition sequences can also be classified by the number of bases in its recognition site.

Many of them are palindromic, meaning the base sequence reads the same backwards and forwards (rotationally symmetrical, the sequence in the 5'to 3 direction on one DNA strand is the same as the sequence in the 5'to 3' direction on the other strand.) and cutting of these recognition sequences in DNA differ by two way, the first called sticky-end "overhang" that producing differences in the length, sequence and strand orientation (5' end or 3' end) such as EcoRI recognizes the palindromic sequence GAATTC and cuts between the G and the A on both the top and bottom strands, leaving an overhang called sticky ends.

5- GAATTC-5
5-CTTAAG-5

Some restriction enzymes cut both DNA strands at the same site in at opposite base, called a blunt end such as Smal.

5 – CCCGGG- 5
5 –GGGCCC-5

The value of restriction endonucleases lies in their specificity. Each particular enzyme recognizes a specific sequence of bases in the recognition sequences being four, six, or eight base pairs in length. Thus, given that there are four bases in the DNA, and assuming a random

distribution of bases, the expected frequency of any particular sequence can be calculated as 4" where n is the length of the recognition sequence. This predicts that tetranucleotide sites will occur every 256 bp, hexanucleotide sites every 4096 bp and octanucleotide sites every 65536 bp. Different restriction enzymes generate different ranges of DNA fragment sizes. Thus, an enzyme recognizing a tetranucleotide sequence will produce shorter DNA fragments than hexa and octanucleotide sequence recognizing enzymes.

Types of Restriction enzymes

Restriction endonucleases are four types (I, II, III, and IV) classified according to their structure, recognition site, cleavage site and cofactor requirements. However, amino acid sequencing revealed that at the molecular level there are many more than four different types. Most of the enzymes commonly used today are type II enzymes, which have the simplest mode of action

Type I restriction enzymes

-Require ATP, S-adenosylmethionine (SAM) and magnesium ions (Mg2+) for their activity.
- Cleaves DNA as far as 1000 bp firom the recognition site.
-Usually large enzymes with many subunits.
Ecok is a type I endonuclease cleaves DNA in a random fashion away from the recognition site described in 1968.

Type II restriction enzymes.

- Requires only Mg2+.
- Cleaves DNA at defined positions close to or within its recognition sequence. Type II enzyme has a number of advantages over type I and III enzymes which makes it the most commonly used in biotechnology include:
-Restriction and modification are mediated by separate enzymes so it is possible to cleave DNA in the absence of modification.
- The restriction activities do not require cofactors such as ATP or SAM, making them easier to use.
- Most important of all, type II enzymes recognize a defined, usually symmetrical, sequence and cut withinit.
-They are typically smaller than Type 1 and 3.
The HindII, HindIII and EcoRI restriction enzymes are all examples of type II enzymes.

Type III restriction enzymes

-Requires Mg2t, ATP and is stimulated by SAM.

-They recognize short, non-palindromic sequences and cleaves outside its recognition site approximately 25 bp from this site.
-Very large with many subunits.

Type III restriction enzymes examples are EcoP1 and EcoP15.

Type IV restriction enzymes

-Requires Mg2+ and SAM

-Cleaving both DAN strands on both sides of its recognition

Some restriction enzymes

Enzyme	Source organism	Restriction recognition site in double-stranded DNA	Structure of the cleaved products
(a)			
EcoRI	*Escherichia coli*	5′ —G—A—A—T—T—C— / —C—T—T—A—A—G— 5′	—G / 5′ A—A—T—T—C— / —C—T—T—A—A 5′ / G— — 5′ overhang
PstI	*Providencia stuartii*	5′ —C—T—G—C—A—G— / —G—A—C—G—T—C— 5′	—C—T—G—C—A 3′ / G— / —G / 3′ A—C—G—T—C— — 3′ overhang
SmaI	*Serratia marcescens*	5′ —C—C—C—G—G—G— / —G—G—G—C—C—C— 5′	—C—C—C / G—G—G / —G—G—G / C—C—C — Blunt ends
(b)			
HaeIII	*Haemophilus aegyptius*	5′ —G—G—C—C— / —C—C—G—G— 5′	—G—G / 5′ C—C— / —C—C 5′ / G—G — Blunt ends
HpaII	*Haemophilus parainfluenzae*	5′ —C—C—G—G— / —G—G—C—C— 5′	—C / C—G—G / —G—G—C 5′ / C— — 5′ overhang

Application

The ability of restriction enzymes to reproducibly cut DNA at specific sequences has led to the widespread use of these tools in many molecular biology techniques.

Lecture 3

Cloning Vectors

A cloning vector is a small piece of DNA, taken from a virus, a plasmid, or a cell of higher organism, that can be stably maintained in an organism, and into which a foreign DNA fragment can be inserted for cloning purposes.The vector therefore contains features that allow for the convenient insertion or removal of a DNA fragment to or from vector,

Cloning is generally first performed using Escherichiacoli, and cloning vectors in E. coli include plasmids, bacteriophages (such as phage λ), cosmids, and bacterial artificial chromosomes (BACs), while cloning vectors in yeast include yeast artificial chromosomes (YACs).

The host/ vector system is important to molecular biology.The vast majority of molecular cloning experiments utilize the bacterium*Escherichia coli* for the propagation of cloned DNA fragments. The advantages of *E. coli* that acceptance as the genetic engineering organism of choice are:

- It is easy to grow in simple, inexpensive growth medium.
- The organism has a rapid doubling time of about 20–30 minutes.
- Laboratory strains of *E. coli* are generally safe.
- Its genetics are well understood and it has a fully mapped and sequenced genome.
- Extra-chromosomal copies of DNA (plasmids and bacteriophage DNA) canbe used to carry foreign DNA fragments.

The ability to propagate DNA in a host cell requires a **vector** defined as autonomously replicating DNA molecules that can be used to carry foreign DNA fragment.

Main characteristics of the vector:-

A vector must possess the following characteristics to make it useful formolecular cloning:

- The ability to self-replicate (have origin of replication).
- Contain at least one or more restrictionsites for the insertion of the DNA fragments wants to be cloned.
- A selectable marker is a gene coding desirable feature used for identification of transformed from non-transformed cells. These markers may be antibiotic resistance genes such as ampicillin and tetracycline resistance gene.Or may be gene product that are give fluorescence or color reactions whenreact with substratesuch aslacZgene encode for B-galactosidasewhen react withX-Gal present in media give colored colonies (blue for non-transformed cells/ whitefor transformed cells).

Small size of the vector is more desirable that increase the efficiency of transformation and easy to manipulate.

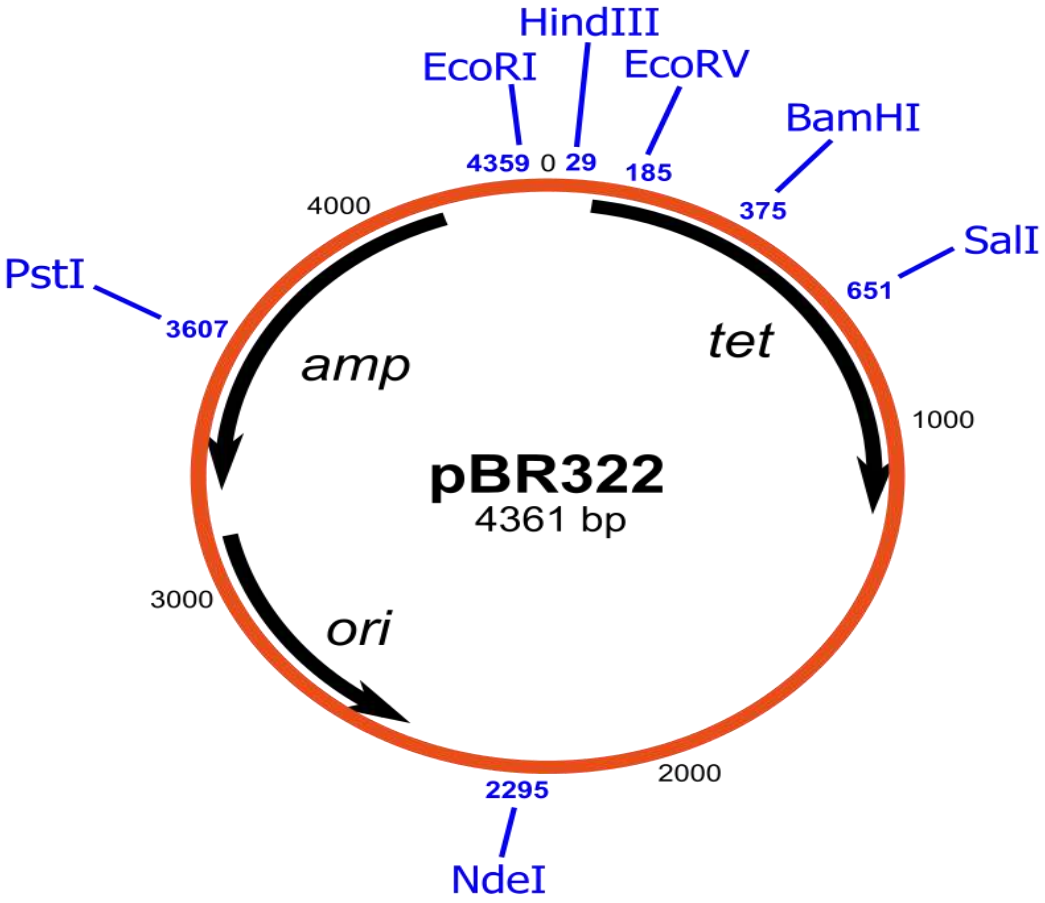
Vectors that commonly used in cloning experiments are:

1. **Plasmids**

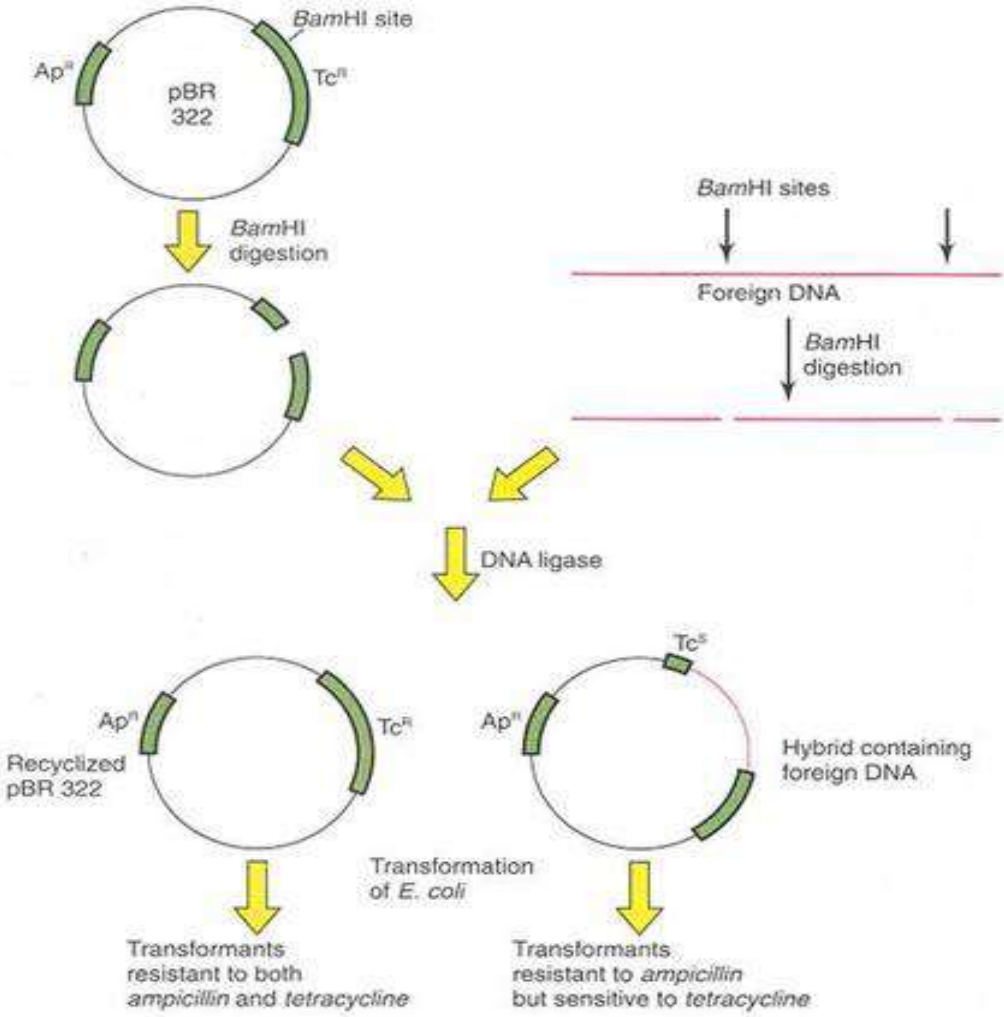
Plasmids are naturally occurring extra-chromosomal DNA fragments that are stably inherited from one generation to another. Plasmids are carrying a gene that encodes resistance to antibiotics, certain toxins or heavy metals, or thatproduces DNA restriction and modification enzymes.The copy number of the plasmids can vary from 1-2 or multiple copies (10-200)/cell. Plasmids are used to clone small pieces of DNA no more than 10 Kb in a site called **polylinker**is a short segment of DNA which contains many restriction sitesthat plasmid can cut with any of restriction enzymes and the desirable DNA can then be ligated into this site.

pBR322

The plasmid pBR322 was the first and commonly used vector in *E. coli* created in 1977.It is a small plasmid (4361 bp) contain15-20 copy number / cell which can be increased 200-fold by chloramphenicol amplification, also pBR322 contains origin of replication, and it has two antibiotic resistance genes to ampicillin and tetracycline. The plasmid has unique restriction sites for more than 40restriction enzymes.



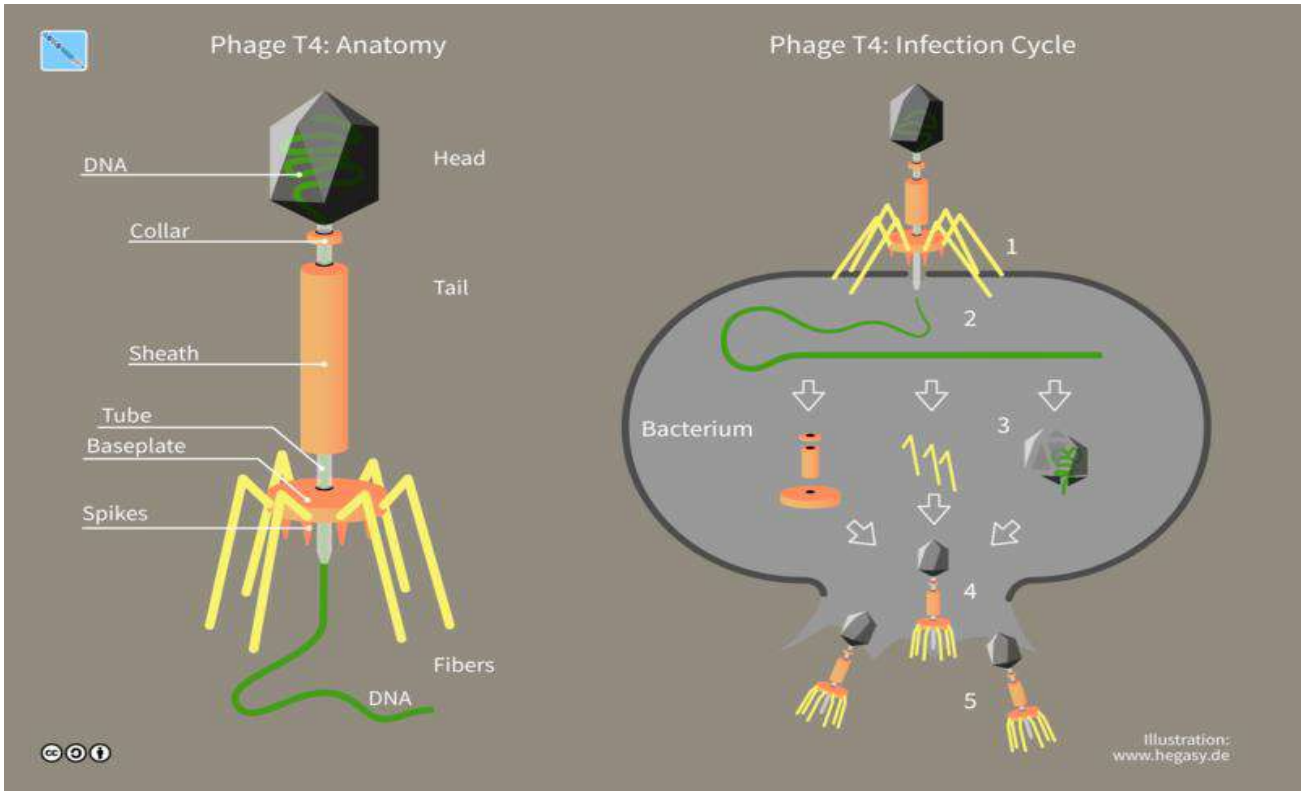
The antibiotic resistance genes in pBR322 allow for selection ofrecombinants (transformed cells) in a process called **insertional inactivation**. If wewant to clone a DNA fragment into the BamHI site of pBR322, theinsert DNA will interrupt the gene (nonfunctional) responsible for tetracycline resistance, butthe gene for ampicillin resistance will not be altered. Transformed cells aregrown on medium containing both ampicillin and tetracycline. The cellsthat die due to presence of tetracycline select asrecombinantscontain the foreign DNA fragment.


BamHI site
Ap^R
Tc^R
pBR
322
BamHI
digestion
BamHI sites
Foreign DNA
BamHI
digestion
DNA ligase
Tc^S
Ap^R
Tc^R
Recyclized
pBR 322
Ap^R
Hybrid containing
foreign DNA
Transformation
of E. coli
Transformants
resistant to both
ampicillin and tetracycline
Transformants
resistant to ampicillin
but sensitive to tetracycline

Lecture 4

2. **Bacteriophages**

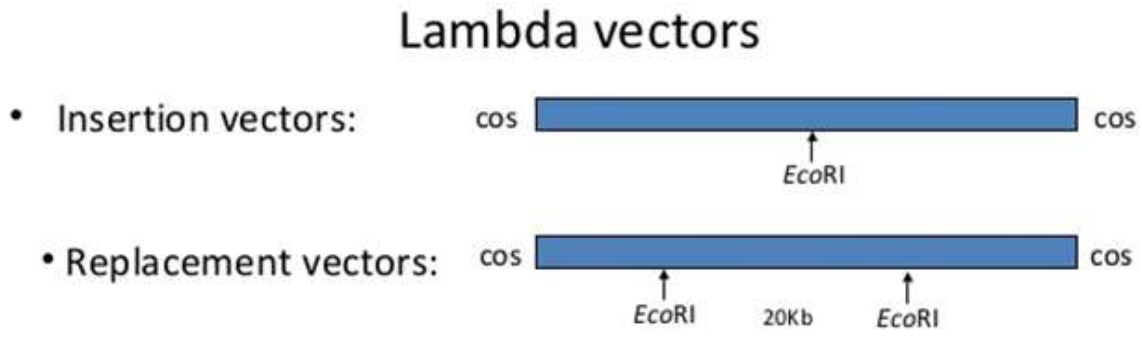
A bacteriophage, also known as a phage, is a virus that infects and replicates within bacteria and archaea.Phages are vectors larger than plasmids can inserts up to 40 Kb.**lambdaphage (λ)**, is bacteriophage that infects the bacterial species *E. coli*. This virus has a temperate lifecycle that allows it to either reside within the genome of its host through lysogeny or enter into a lytic phase. λDNA has origin of replication and contains few unique restriction sitesinto which foreign DNA fragments could be cloned.



Two basic types of λ vector are:

Insertional vector:foreign DNA (8-10 Kb) is inserted into specific restriction site and can be introduced without removal of the stufferfragment. These are useful in cloning of small DNA fragments such ad cDNA therefore used in cDNA libraries.

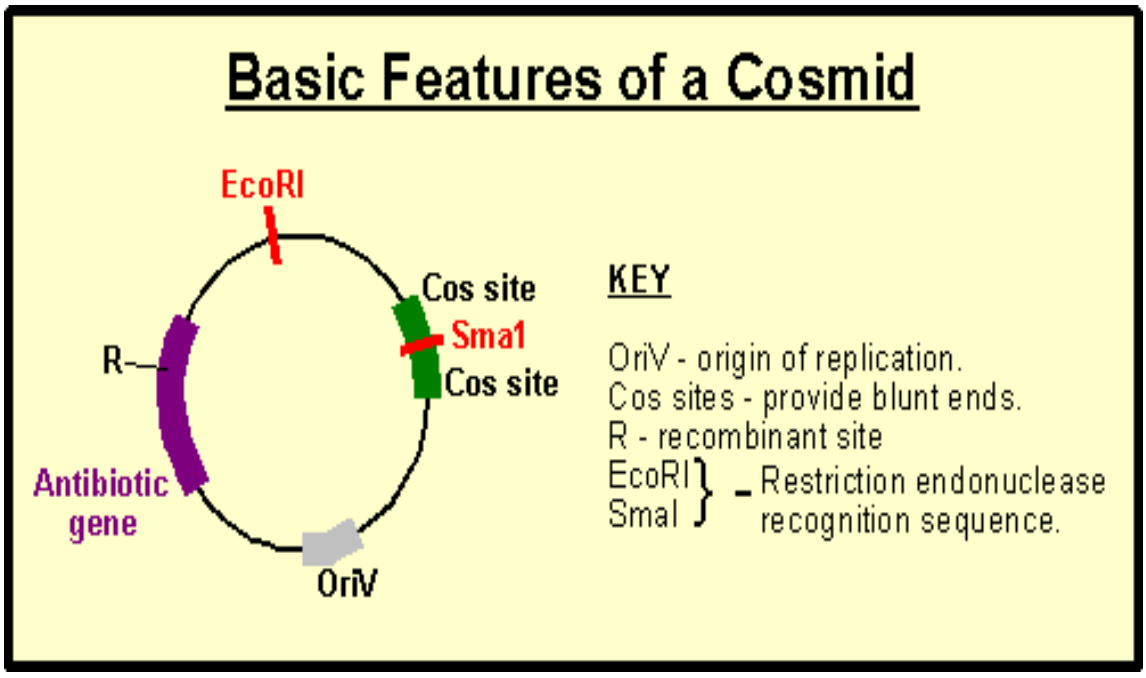
Replacement vector: foreign DNA replaces the stuffer fragmentof vector.Replacement vectors are used for cloning of large DNA fragment(about 20 Kb)and are used in genomic libraries of higher eukaryotes.



3. Cosmids

A cosmid is a type of hybrid plasmid that contains a Lambda phage cos sequence. Cosmids (cos sites + plasmid = Cosmids) DNA sequences are originally from the lambda phage. They are often used as a cloning vector in genetic engineering. Cosmids can be used to build genomic libraries. They contain 37 to 52 (normally 45) kb of DNA, limits based on the normal bacteriophage packaging size.

As plasmids, cosmids contain an origin ofreplication, a selectable marker and a unique restrictionsite into which DNA fragments can be ligated. *cos* site is essential for packaging of cosmid DNA into phage particles.

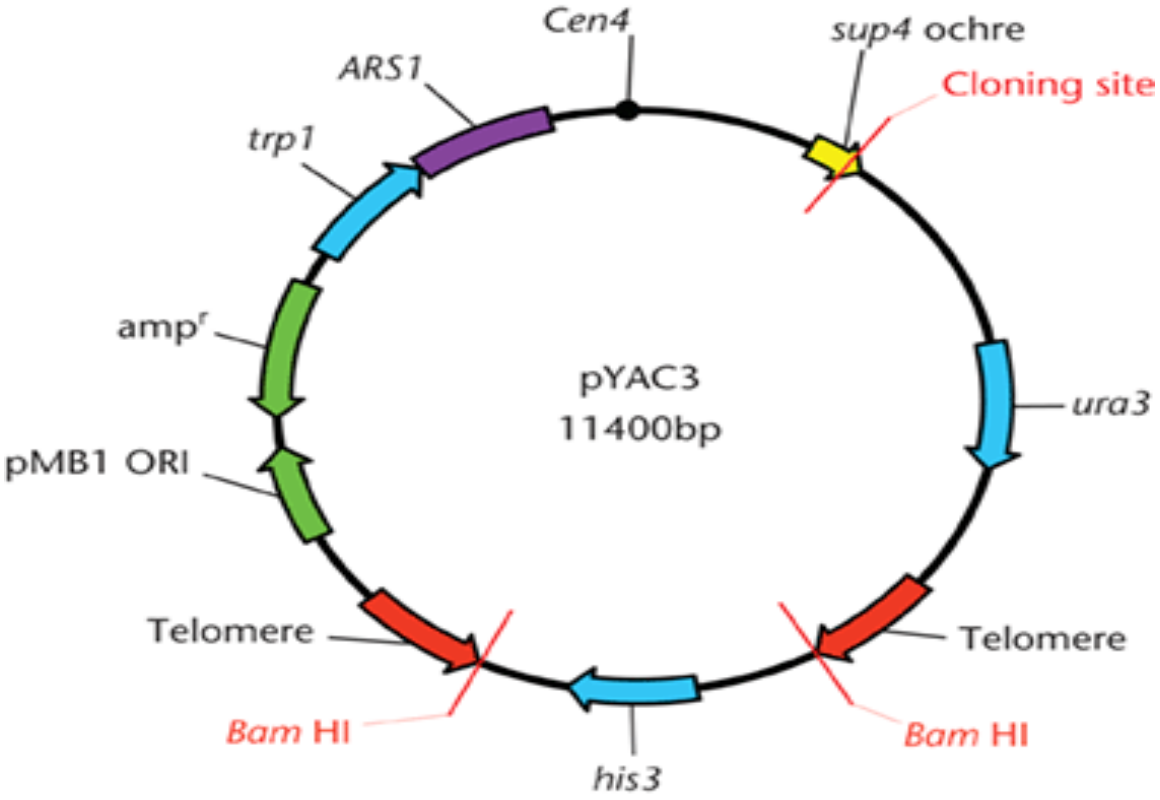


Artificial chromosome

1. **YACs**

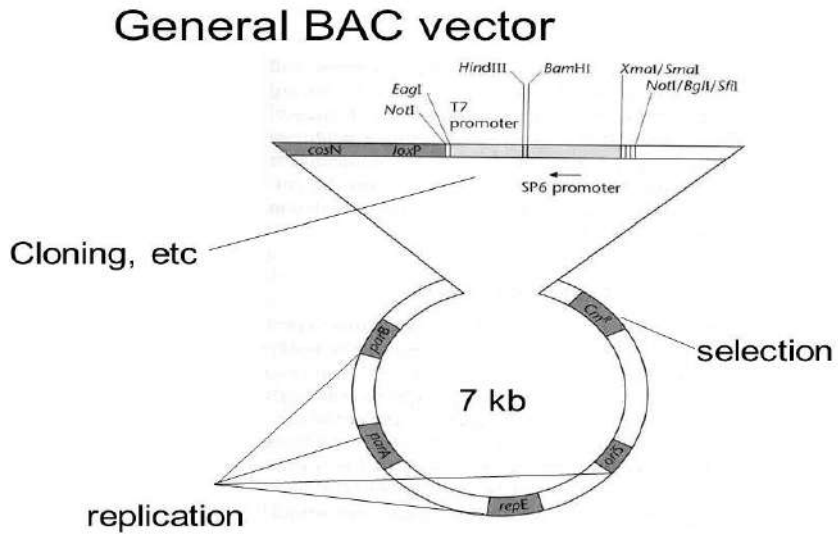
   Yeast artificial chromosomes (YACs) are genetically engineered chromosomes derived from the DNA of the yeast, Saccharomyces cerevisiae,which is then ligated into a bacterial plasmid. By inserting large fragments of DNA, from 100–1000 kb.

In order to propagate the vector in bacterial cells prior to insertion of genomic DNA, YAC contain a bacterial replication origin andbacterial selectable marker such as ampicillin resistance gene. There are difficulties associated with working with YACs such asvery large DNA molecules are very fragile and tend to breakage, leadingto problems of rearrangement.



2. **BACs**

   A bacterial artificial chromosome (BAC) is a DNA construct, based on a functional fertility plasmid (or F-plasmid), used for transforming and cloning in bacteria, usually E. coli. BACs are capable of carrying 150- 350 Kb of inserted DNA sequence.



BACs contain origin of replication sequencederived from*E. coli* plasmid and four F-factor genes required for replication and maintenance ofcopy number, also contain multiple unique restriction sites and a selectableantibiotic resistance makers. Additionally, the BAC contains a λ cos site that used for specific cleavage during restriction mapping.

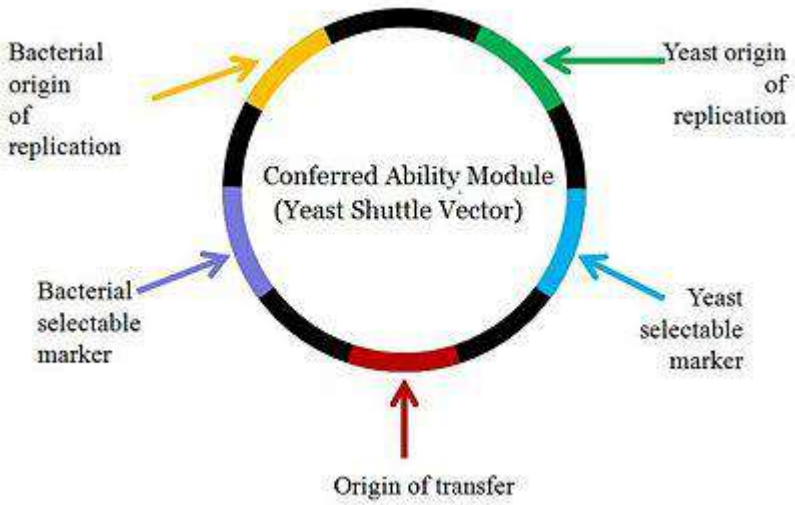
The DNA inserted into a BAC appears to be very stable. It can survivefor many hundreds of generations in *E. coli* cells, and appears to be lesslikely to rearrangements and deletions.

Vector	**host**	**Insert size**
plasmid	*E. coli*	10 kb
λ phage	*E. coli*	5–25 kb
*λ*cosmids	*E. coli*	35–45 kb
BACs	*E. coli*	≤300 kb
YACs	*Saccharomyces cerevisiae*	200–2000 kb

Animal and Plant Vectors (Shuttle vectors)

There are some vectors developed for transforming plant and animal cells and used in eukaryotic systems. These are often called shuttle vectors as they replicate in both prokaryotic and eukaryotic hosts. The common features of these vectors are:

- They are capable of replicating in two or more types of hosts including prokaryotic and eukaryotic cells such as bacteria and yeast.
- They replicate autonomously or integrate into the host genome and replicate when the host cell multiplies.
- Commonly used for transporting genes from one organism to another (transforming animal and plant cells).


Bacterial origin of replication
Yeast origin of replication
Conferred Ability Module (Yeast Shuttle Vector)
Bacterial selectable marker
Yeast selectable marker
Origin of transfer

Lecture 5

Nucleic acid hybridization

In molecular biology, hybridization is a phenomenon in which single-stranded deoxyribonucleic acid (DNA) or ribonucleic acid (RNA) molecules anneal to complementary DNA or RNA through base pairing by the formation of specific hydrogen bonds between nucleotide bases. Hybridization is a part of many important laboratory techniques such as polymerase chain reaction (PCR), northern blots and Southern blotting, and DNA sequencing.

Hybridization in molecular methods uses single-stranded nucleic acid probes of defined sequence to hybridize a target DNA or RNA of interest. The probe can be labeled with enzymes, chemiluminscent, radioisotopic, magnetic particles or fluorescent moieties that can be readily detected or captured by automated instruments.

Denaturation

DNA is a two stranded molecule, held together by hydrogen bonds that separated under appropriate conditions during replication of DNA and the synthesis of RNA. Experimentally, boiling a solution of DNA adds energy and breaks these bonds, making the DNA single-stranded. This is known as denaturationof the DNA (melting).

The melting process can be monitored by the difference in absorbance between double stranded and single stranded at wave length at 260nm. When the temperature of a DNA solution is raised, the absorbance of duplex structure remains constant until it is melt to their component (single strand) resulting in rapid increase of absorbance.

The melting temperature (Tm) is defined as the temperature at which 50% of all molecules of a given DNA sequence are double strand, and 50% are present as single strands.

The Tm is affected by a number of factors:

1. **Concentration, length and sequence of DNA**. In long pieces and high concentration of DNA, the more GC bonds present therefore higher temperature is required. The nucleotide pair A=T has a weaker bond than the G≡C. Therefore G≡C are more resistant to denaturation lead to Tm increase linearly with the proportion of GC pairs in the DNA. Nucleotides on the same strand can interact with each other, forming so-called secondary structures that increase the Tm.

2. **Concentration of ions in the solution**, most notably Mg+ and K+ (The concentration of ions affects Tm because DNA is electrically charged.)

Tm= 4(G+C) + 2(A+T)

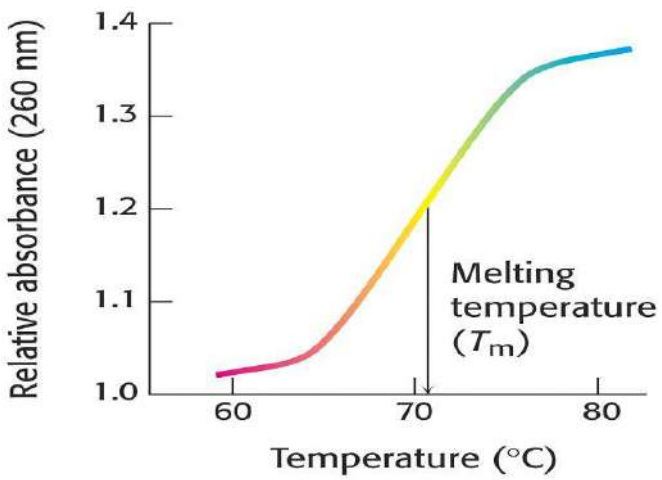


Figure 1: Denaturation of DNA describe Tm

Annealing

The term is often used to describe the binding of a DNA probeor primer to a DNA strand during a polymerase chain reaction. The term is also often used to describe the reformation (**renaturation**) of complementary strands by hydrogen bonds that were separated by heat.

Types of Nucleic acid hybridization

DNA–DNA hybridizationrefers to the degree of genetic similarity (genetic distance) between DNA sequences. When several species are compared that way, the similarity values allow the species to be arranged in a phylogenetic tree.

The DNA of one organism is labeled, and then mixed with the unlabeled DNA to be compared against. The mixture is incubated to allow DNA strands to dissociate and re-anneal, forming hybrid double-stranded DNA. Hybridized sequences with a high degree of similarity will bind more firmly, and require more energy to separate them than dissimilar sequences.

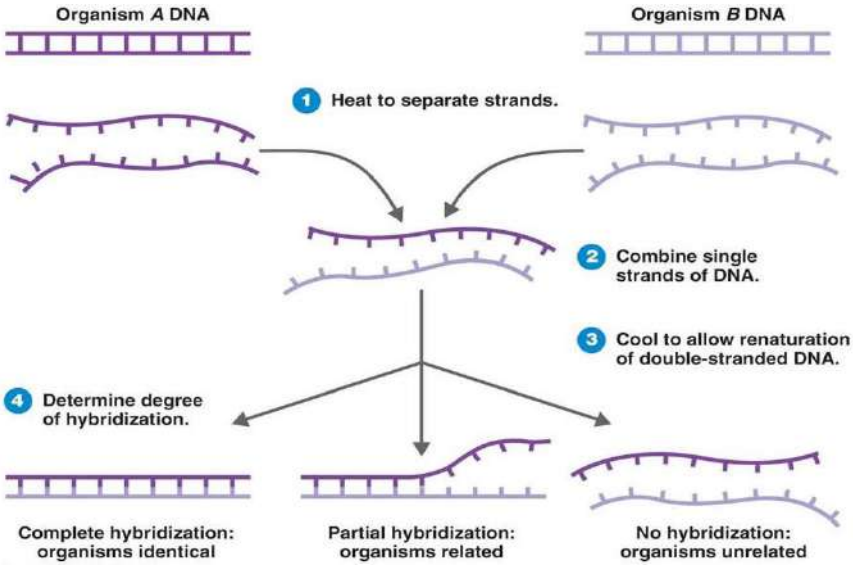


Figure 2: DNA–DNA hybridization

DNA-RNA Hybridization

DNA-RNA hybridization follows the same base pairing rules as two complementary DNA strands and the most common source of DNA complementary to an mRNA. In DNA-RNA hybrid formation,since RNA is single-stranded, no denaturation is necessary if RNA is the target nucleic acid. G base pairs with C, A of the RNA pairs with T of the DNA, and U of the RNA pairs with A of the DNA.

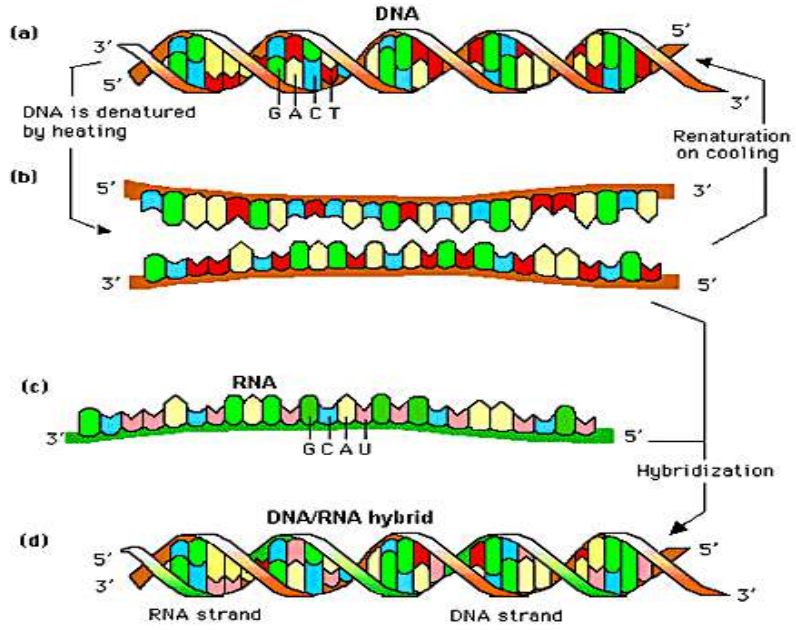


Figure 3: DNA–RNA hybridization

Liquid and solid phaseHybridization (based on the type of reaction)

Hybridization may be performed with both the probe and test sequences in the liquid phase or with test sequences bound to a solid phase, usually nitrocellulose or nylon membrane. Both methods may be used to quantify the amount of the test sequences present, but solid phase hybridization is also used to locate the position of sequences immobilized on the membrane.

Solid (Sandwich) hybridization is a simple test but has the disadvantage of a slow reaction rate while liquid hybridization methods are fast and easy to perform.

In solid hybridization, Nitrocellulose can bind high molecular weight DNA with high efficiency but below 500bp with poor efficiency. It is fragile have hydrophopic binding and not easily re-probed again. In contrast, nylon membrane can bind nucleic acid smaller than 500bp with high efficiency. It is durable have covalent binding and can be re-probed again several times.

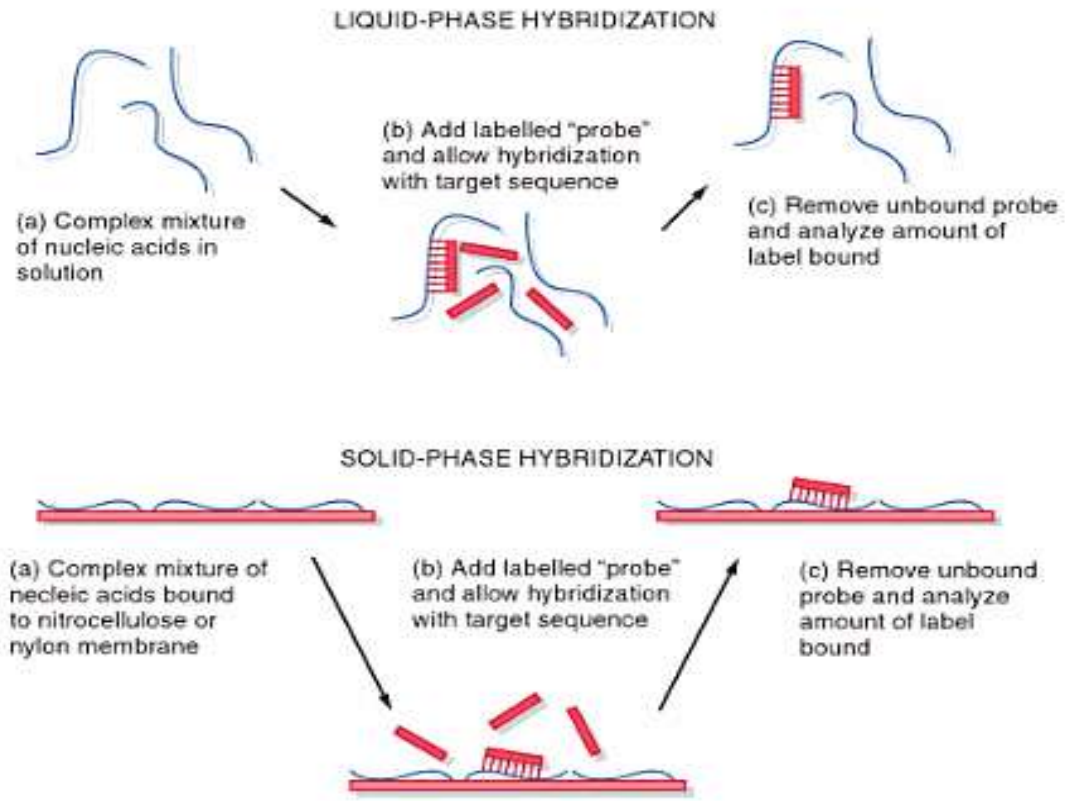


Figure 4: Solid and liquid hybridization

Hybridization Techniques

Many techniques used labeling and hybridization of nucleic acid and other biological molecules on membranes as a key for identification of a fragment of nucleic acid contains a specific

nucleotide sequences that represent a single gene or larger piece of DNA. These techniques include:

Southern Hybridization

The southern hybridization technique helps to identify a fragment of DNA that contains a specific nucleotide sequence.

Hybridization step

1- Isolation and Purification of DNA

Isolation of DNA is done by incubating the specimens with detergents for cell lysis then parts are enzymatically degraded by proteinases.DNA is purified from contaminant by alcohol precipitation and then suspended in buffer for storage.

Lecture 6

2- DNA Fragmentation

Isolated DNAis cleaved with restriction endonucleases (they cleave at specific sites) into fragments of various sizes.

3- Gel Electrophoresis

Gel electrophoresis is a method for separation and analysis of macromolecules (DNA, RNA and proteins) and their fragments, based on their size and charge.

Nucleic acid molecules are separated by applying an electric field to move the negatively charged molecules through a matrix of agarose or other substances. Shorter molecules move faster and migrate farther than longer ones because shorter molecules migrate more easily through the pores of the gel. This phenomenon is called sieving.

Fragments of DNA are easily separated by gel electrophoresis basis of their sizes.

If some of the DNA fragments are larger than 15 kb, the gel may be treated with an acid (dilute HCl) which de-purinates the DNA fragments, breaking the DNA into smaller pieces, thus allowing more efficient transfer from the gel to membrane.

4- Denaturation of Separated DNA

DNA bands in gel are denatured with an alkaline solution such as NaOH. An alkaline environment may improve binding of the negatively charged thymine residues of DNA to a positively charged amino groups of membrane, separating it into single DNA strands and destroys any residual RNA that may still be present in the DNA.

DNA is then neutralised with NaCl or tris buffer to prevent rehybridization.

5- Blotting

Is a method of transferring proteins, DNA or RNA, onto a carrier (for example, a nitrocellulose, or nylon membrane), this is done after a gel electrophoresis, transferring the molecules from the gel onto the blotting membrane. After blotting, the transferred proteins, DNA or RNA are then visualized by colorant staining (for example, silver staining of proteins).

The DNA is transfer from gel to membranes by:

- Capillary action of buffer from bottom filter paper through the gel carrying denatured DNA present in gel. This process takes several hours or overnight.
- Rapid elecrophoretic transfer of DNA from gel.
- Vacum driven blotting is rapid transfer and prevent DNA diffusion in gel.

1. The nitrocellulose membrane is removed from the blotting stack and DNA is permanently immobillized on the membrane by: baking at 80°C for 2hrs.or by U.V to permanently attach the transferring DNA to membrane.

6- Hybridization with radioactive probes

The labeled probe is added to the blocked membrane in buffer and incubated for several hours to allow the probe molecules to hybridize with the target DNA.To ensure the specific binding of the probe to the sample DNA, most common hybridization methods used deionized formamide and detergents such as SDS to reduce non-specific binding of the probe.

7- Washing

After hybridization, excess probe is washed from the membrane and the pattern of hybridization is visualized on X-ray film by autoradiography in the case of a radioactive or fluorescent probe, or by development of color on the membrane if a chromogenic detection method is used.

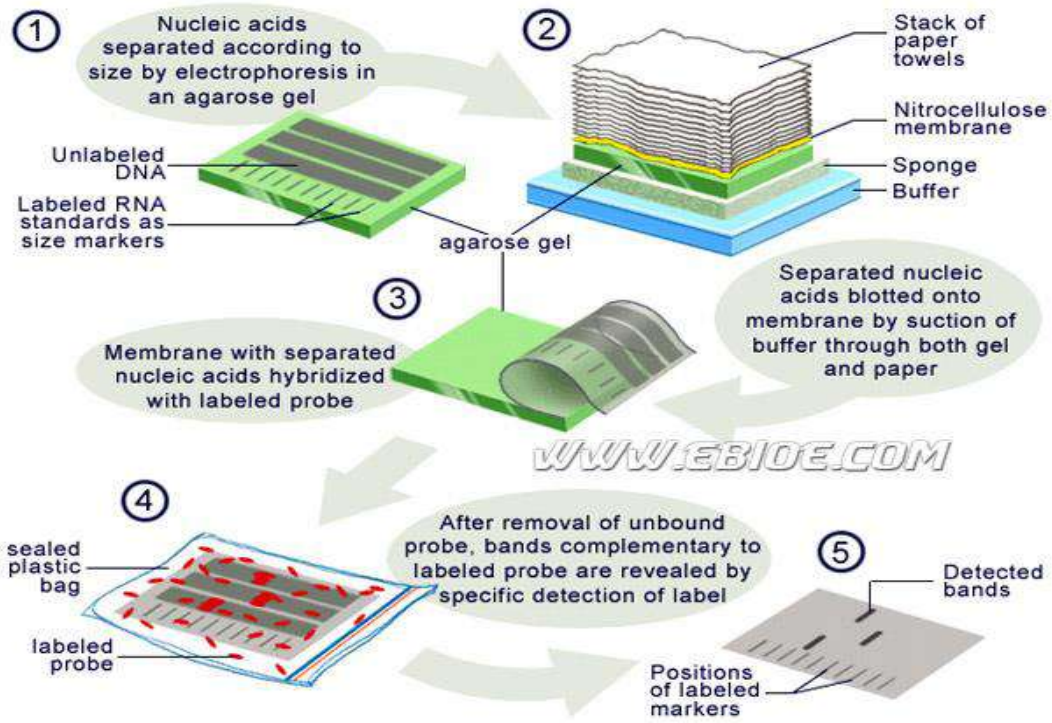


Figure 5: Southern Blotting

Application

- Identify specific DNA in DNA sample or isolate a desired DNA segment for recombinant DNA technology

- Identify mutations, deletions, and gene rearrangements.
- Prognosis of cancer, prenatal diagnosis of genetic diseases.
- Gene mapping.
- Used in phylogenetic analysis

Northern hybridization

Detection of RNA sequences in a sample using probes. RNA is separated on gel and transferred to a membrane. Specific RNA molecules are then detected byhybridization using labelled single-stranded DNA sequences (RNA-DNA hybridization). It's not need step of denaturation.

The advantages and disadvantages of using northern blotting include:

- The detection of RNA size and the membranes can be stored and reprobed for years after blotting.
- Detect small changes in gene expression for small number of genes.
- A problem in northern blotting is often sample degradation by RNases which can be avoided by sterilization of glassware and the use of RNase inhibitors.

Colony blot hybridization

Colony blot hybridization is applied to DNA or RNA released from blotted microbial colonies. Colony blot hybridization can be used for screening clones or bacterial isolates. It is allowed screening of a large number of bacterial colonies from a single plate.

The steps of this type of hybridization are:

- The microbial colonies are transferred (blotted) to a membrane.
- Nitrocellulose or nylon membrane is placed on top of plate to produce a replica.
- The replica is treated with alkali (NaOH) to initiate both bacterial cell lysis and DNA denaturation then sheet is treated with proteases in order to remove the protein and leave the denatured DNA bound to the membrane.
- The sheet is then baked at 80 ◦C to firmly adherethe DNA to the membrane.
- The single-stranded DNA replica can hybridizedwith a labelled single-stranded probe and visualized by UV or autoradiography to reveal the location of colonies on the original plate.

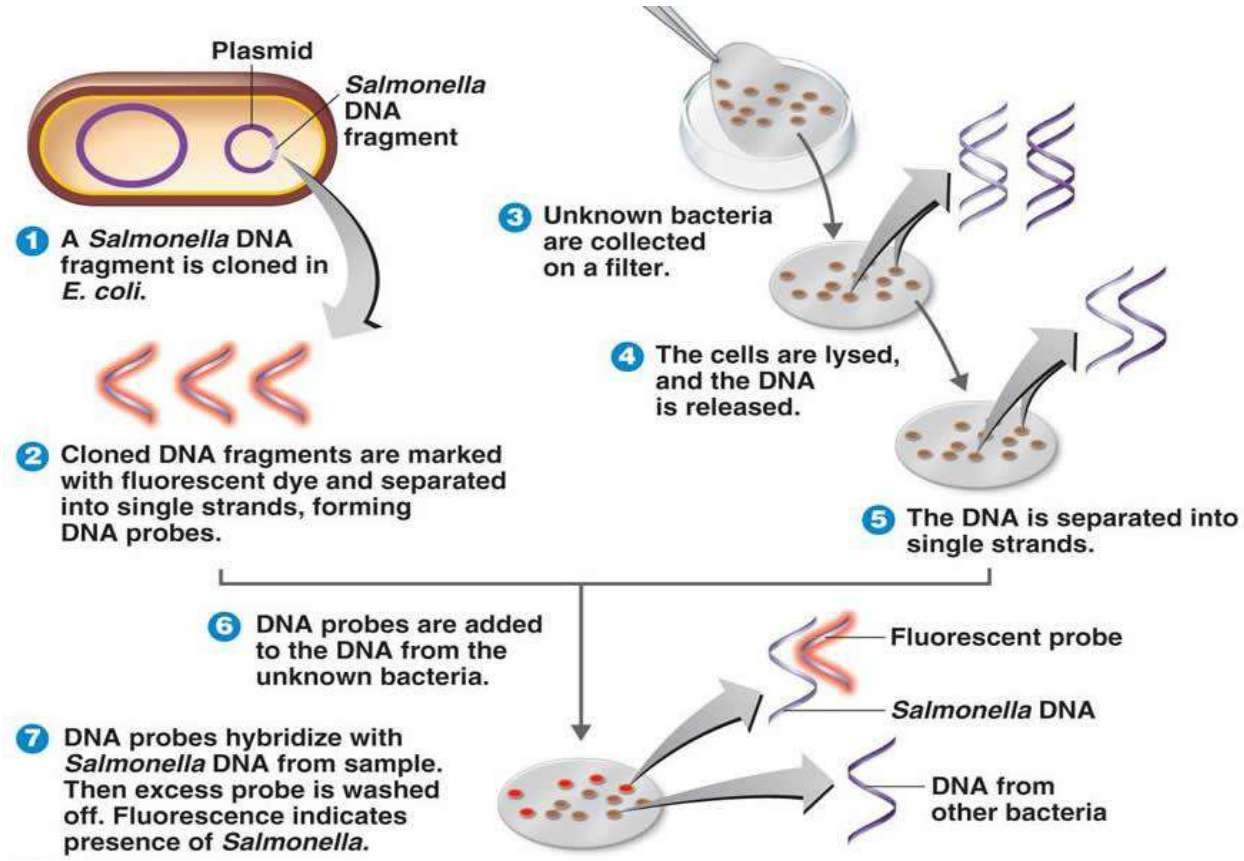


Figure 6: Colony blot hybridization

In situ hybridization

In situ hybridization (ISH) is a type of hybridization that uses a labeled complementary DNA, RNA or modified nucleic acids strand (i.e., probe) to localize a specific DNA or RNA sequence in a portion or section of tissue (in situ).

The term *in situ* hybridization is restricted to whole cell hybridizations in which the cells are detected in their natural microhabitat (in the normal location). Fluorescent in situ hybridization can be used to detect and quantify specific microorganisms while maintaining their morphological integrity, without nucleic acid extraction. Sample cells are fixed with chemicals to increase their permeability and allow the probe to enter the cells. After hybridization the slides air dried and then visualized using fluorescence microscopy.

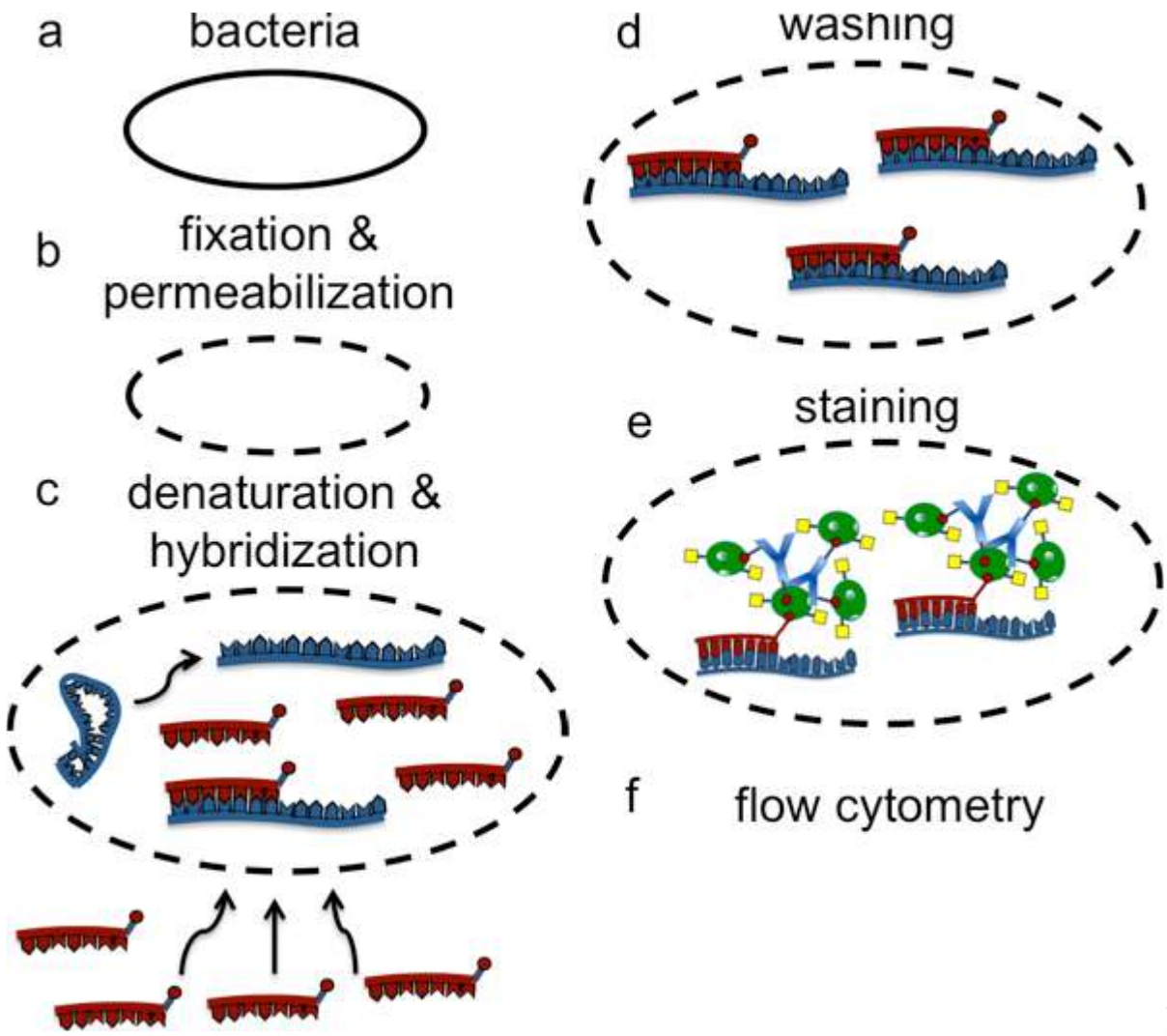


Figure 7: In situ hybridization

Western blotting

The western blot (protein immunoblot) is a widely used analytical technique used in molecular biology, to detect specific proteins in a sample of tissue homogenate or extract.

In brief, the sample undergoes protein denaturation, followed by gel electrophoresis. A synthetic or animal-derived antibody (known as the primary antibody) is created that can recognize and binds to a specific target protein. The electrophoresis membrane is washed in a solution containing the primary antibody, before excess antibody is washed off. A secondary antibody is added which recognizes and binds to the primary antibody. The secondary antibody is visualized

through various methods such as staining, immunofluorescence, and radioactivity, allowing indirect detection of the specific target protein.

Applications

- The confirmatory HIV test employs a western blot to detect anti-HIV antibody in a human serum sample.
- Confirmatory test for Hepatitis B infection.
- A western blot is also used as the definitive test for Bovine spongiform encephalopathy.

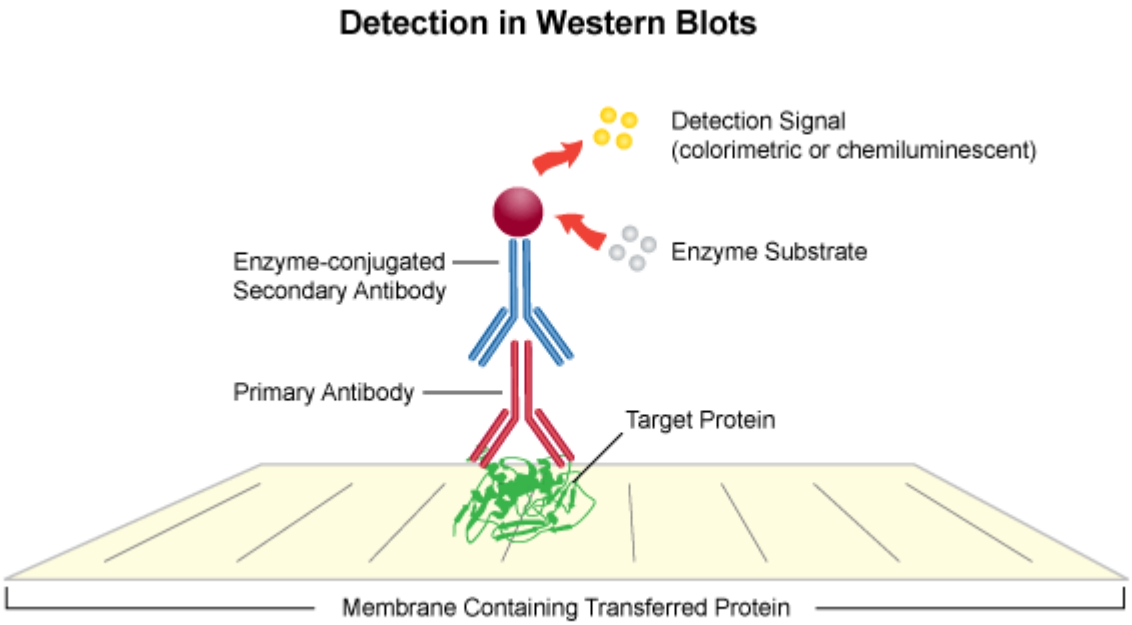


Diagram 2: *Illustration of detection in Western Blots.*

Lecture 7

Recombinant DNA technology

It is a set of molecular techniques for locating, isolating, altering, and studying DNA segments. The term *recombinant* is used because frequently the goal is to combine DNA from two different source that are not found together in nature.

Ex: a human gene might be inserted into a bacterial or viral chromosome. Commonly called **genetic engineering,** Recombinant DNA technology now encompasses an array of molecular techniques that can be used to analyze, alter, and recombine virtually any DNA sequences.

The necessary steps to making recombinant DNA are:

1- Isolating DNA include:
   a- Crude isolation of donor (foreign) DNA.
   b- Crude isolation of plasmid vector DNA.
2- Cutting DNA into large fragments by restriction enzymes (RE) that will recognize specific nucleotide sequences in the DNA and break the DNA chain at those points. Most cut at specific palindromic sites in the DNA (sequence that is the same on both antiparallel DNA strands). These cuts can be generate "sticky or overhanging ends" or a blunt which generate flush ends.
3- Joining DNA: Once have isolated and cut the donor and vector DNAs (plasmid), they must be joined together by DNA ligase. If both have been cut with the same RE, the ends will match up because they are sticky. DNA ligase creates a phosophodiester bond between two DNA ends.

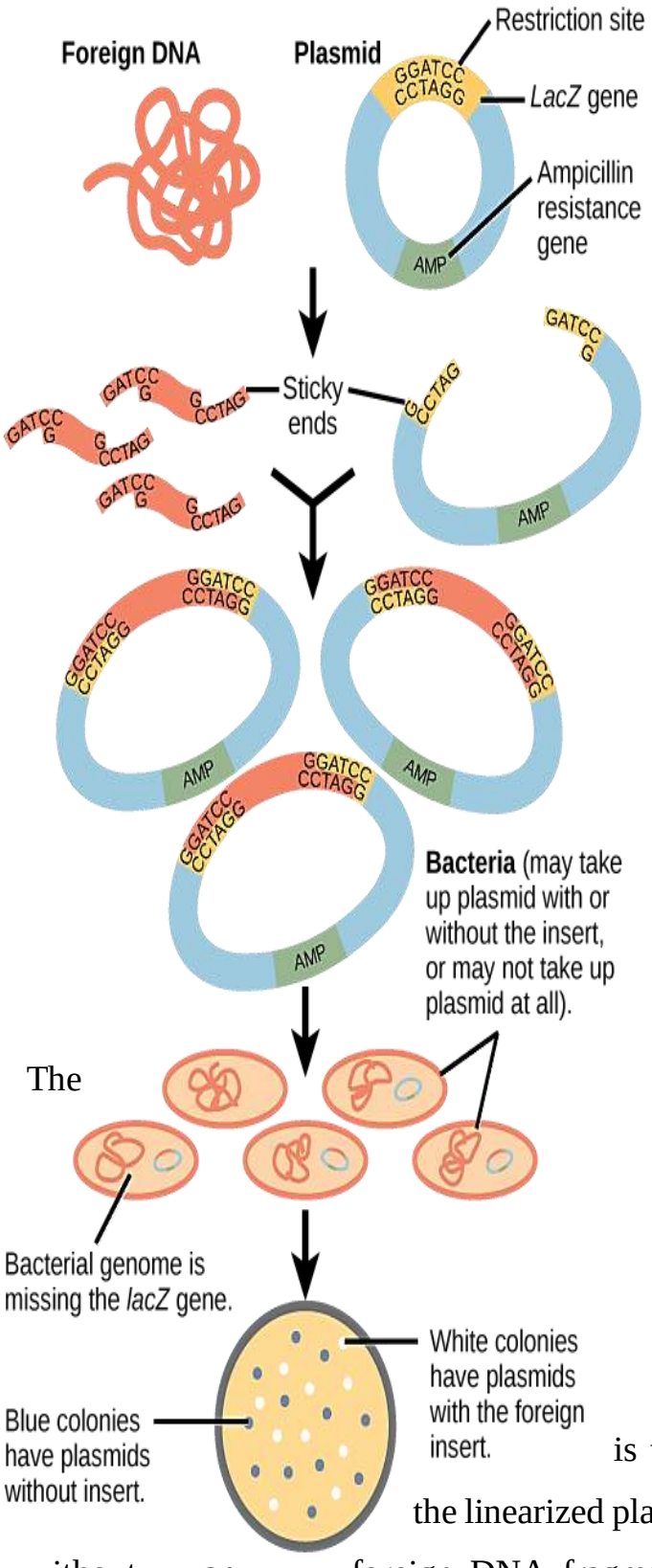
> **Bacterial plasmids** are small, circular DNA molecules that are separate from the rest of the chromosome. They replicate independently of the bacterial chromosome. Useful for cloning DNA inserts less than 20 kb (kilobase pairs). Inserts larger than 20 kb are

4- Amplifying the recombinant DNA to recover large amounts of the recombinant DNA molecule, it must be amplified. This is accomplished by transforming the recombinant DNA into a bacterial host strain. Once in a cell, the recombinant DNA will be replicated. When the

cell divides, the replicated recombinant molecules go to both daughter cells which themselves will divide later. Thus, the DNA is amplified.

5- Selection of bacterial cells containing rDNA. Usually depend on antibiotics resistance that the plasmid confers upon its host cell.

6- Identification of bacterial colonies containing inserted specific DNA fragment.



The foreign DNA and plasmid are cut with the same **restriction enzyme**, which recognizes a particular sequence of DNA called a *restriction site*. The restriction site occurs only once in the plasmid, and is located within the *lacZ* gene, a gene necessary for metabolizing lactose.

The restriction enzyme creates sticky ends that allow the foreign DNA and cloning vector to anneal. An enzyme called ligase glues the annealed fragments together.

The ligated cloning vector is transformed into a bacterial host strain that is ampicillin sensitive and is missing the *lacZ* gene from its genome.

Bacteria are grown on media containing ampicillin and X-gal, a chemical that is metabolized by the same pathway as lactose. The ampicillin kills bacteria without plasmid. Plasmids lacking the foreign insert have an intact *lacZ* gene and are able to metabolize X-gal, releasing a dye that turns the colony blue. Plasmids with an insert have a disrupted *lacZ* gene and produce white colonies.

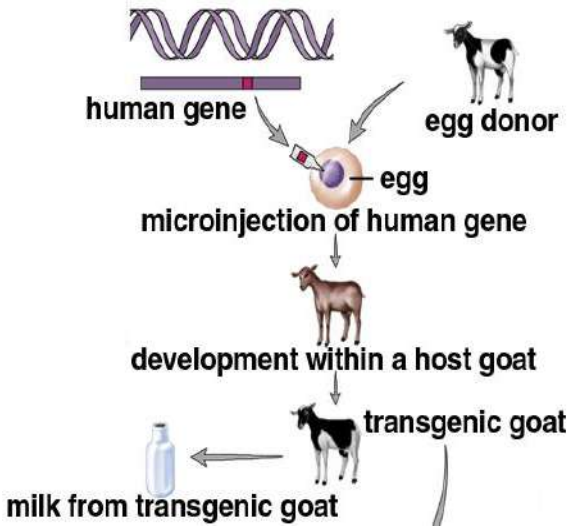
disadvantage of this strategy is that there is the possibility of re-annealing of the sticky ends of the linearized plasmid (vector) leading to formation of the vector itself without any foreign DNA fragment. This problem can be reduced considerably by adopting any one of the following strategies:

- Treating the digested vector with alkaline phosphatase to remove the 5' phosphate group.
- Linearization of the vector (restriction digestion of the vector) and the isolation of the forgoing DNA fragment with two restriction enzymes instead of one (double digestion). So that the two sticky ends of the vector will not be identical, which will prevent the reannealing to some extent.

Applications of recombinant DNA technology:

1. Recombinant DNA technology uses in medicine.
    - For production of vitamins like B12.
    - For production of antibiotics on large-scale.
    - For production blood clotting factor VIII.
    - Recombinant proteins like insulin, enzymes (ex: urikinase, which is used to dissolve blood clots,),human growth hormones (ex: HGH) and production of Interferon.
    - Produce recombinant vaccines in large-scale. Ex. Hepatitis-B vaccine.
    - Used in Gene Therapy:replace the defective genes responsible for hereditary diseases (e.g., haemophilia, phenylketonuria, alkaptonuria) with normal genes.

2. Recombinant DNA technology uses in animal husbandry and sericulture.
    - For enhancement of milk production in cattle.
    - For better meat yield in animals like pigs, cattle, birds.
    - For enhancement of silk production in Seri culture.
    - For better egg yield in poultry birds.
    - For better wool yield from sheep.



3. Recombinant DNA technology uses in agriculture.
    - For biotechnology crops like cotton, vegetables etc. rDNA technology can help produce high yielding plants with desired quality.

- Disease resistant crops like BT-cotton, BT-brinjal are produced to with stand pest attack and there by limit pesticide usage.

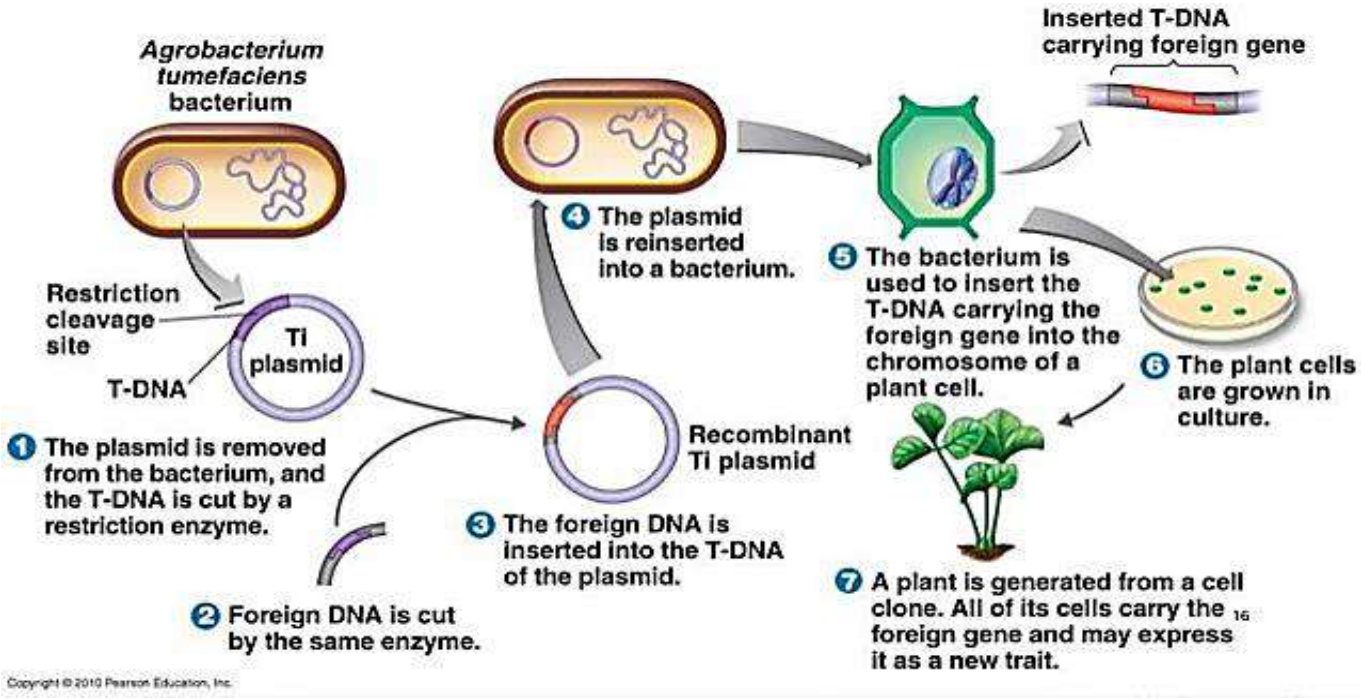


Figure demonstrate the transgenic plant

<u>Along with applications recombinant DNA technology has disadvantages also which are as follows:</u>

1. Antibiotic resistant genes are used as marker gene for identifying transformed cells. This has given rise to **antibiotic resistance human pathogens.**

2. Insect resistant plants produce certain chemicals which are harmful for insects. By repeated growing of such plants causes **resistant insects**. It is possible because exposure to long term may cause mutation in insects.

3. Many **ethical issues** have been raised regarding the use of transgenic animals as animal model and genetically modified food by animal rights organizations.

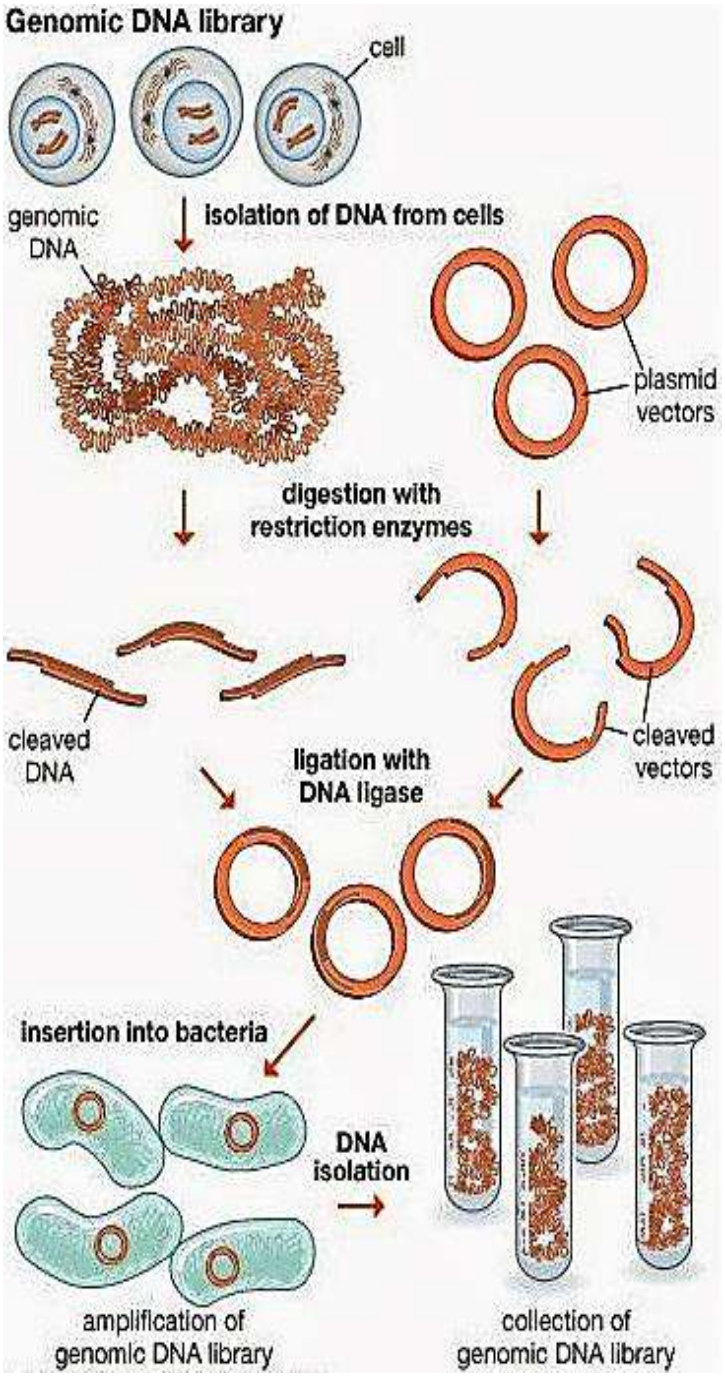
4. Genetically modified organism (plants or animals) can interbreed with natural organism and contaminate natural environment with loss of natural flora and fauna. Also the result of such interbreed is unpredictable and is matter of concern.

5. **Resistance of people** to genetically modified foods. People don't like to eat foods which have been genetically altered.

Lecture 8

Genomic and cDNA Libraries

- Genomic DNA Library:



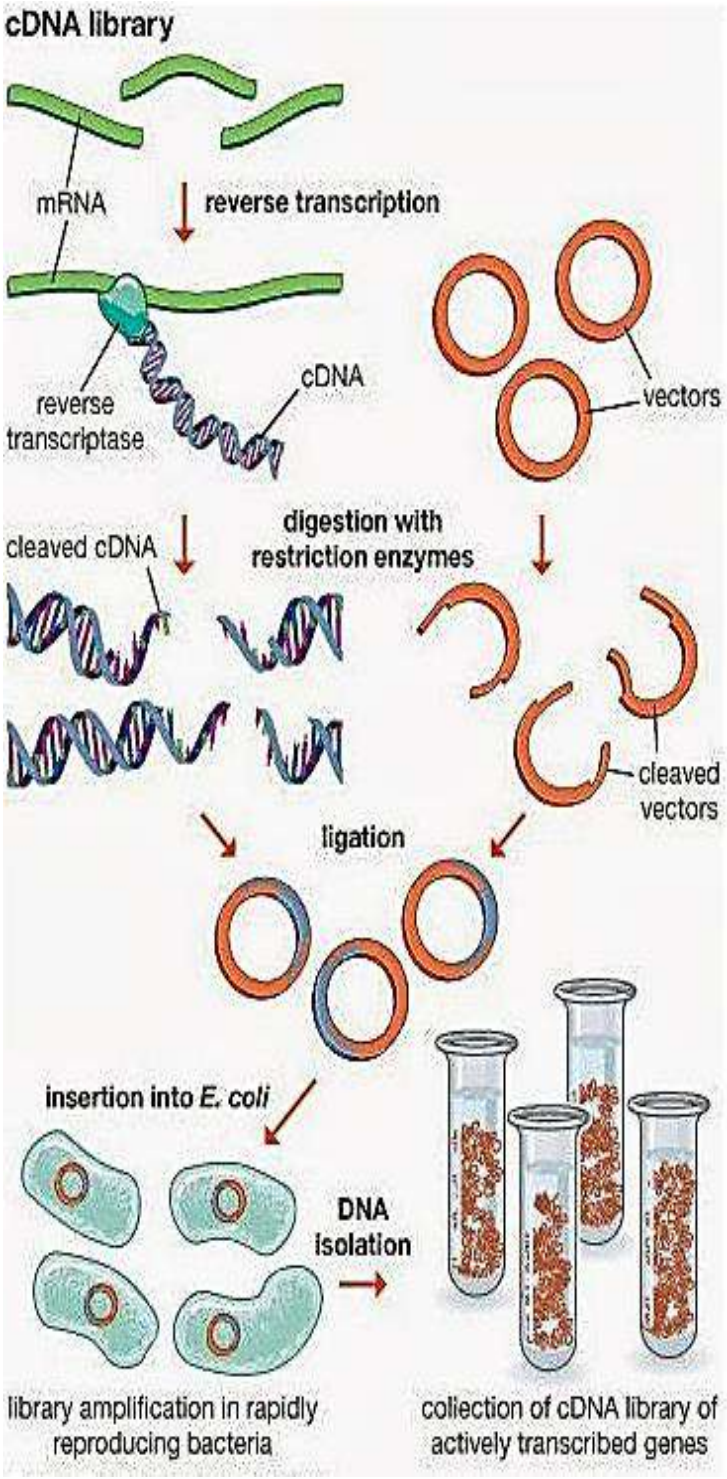
A genomic library is a collection of independently isolated vector linked DNA fragments derived from a single organism. It contains at least one copy of every DNA sequence in the genome. An ideal library is one that represents all of the sequences with smallest possible number of clones.

The genomic DNA libraries can be prepared by the complete digestion of the total genomic DNA with a restriction enzyme into smaller fragments, the fragments are inserted into a suitable vector. The drawback of this method is that sometimes the sequence of interest may contain multiple restriction sites, so digestion with RE results into two or more pieces.

This method ensures that sequences are not excluded from the cloned library simply because of the distribution of restriction sites.In this procedure the randomly fragmented DNA is partially digested with RE which has short recognition sites. The fragments of desired size are collected through agarose gel electrophoresis, so the population of overlapping fragments that are close to random can be cloned directly.

- cDNA Library:

A cDNA library is a combination of cloned cDNA (complementary DNA) fragments inserted into a collection of host cells, which together constitute some portion of the transcriptome (is the set of all RNA molecules in one cell or a population of cells) of the organism and are stored as a "library".



cDNA is produced from fully transcribed mRNA found in the nucleus and therefore contains only the expressed genes of an organism. The mRNAs are used for copying them into cDNAs through the use of reverse transcriptase. Then the cDNA molecule can be made double stranded and cloned.

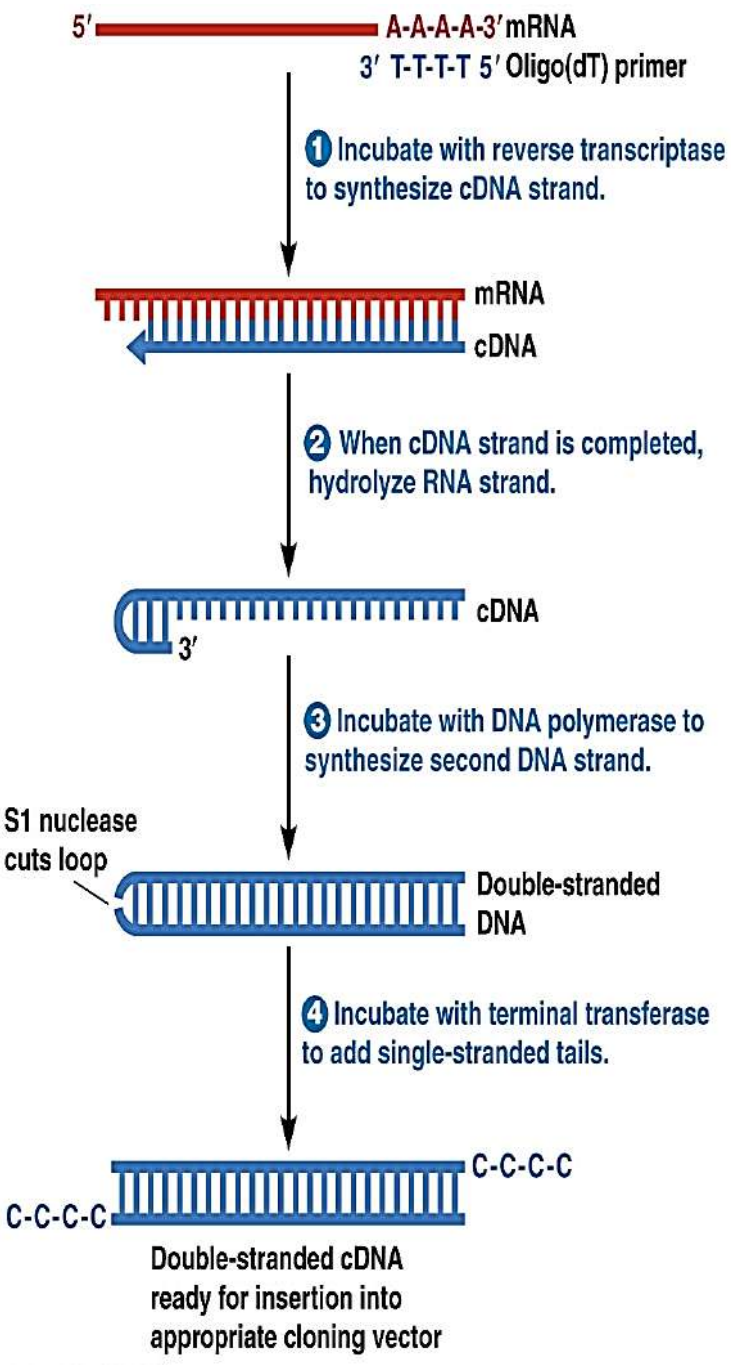
cDNA clones will differ from genomic clones in lacking the introns pre-sent in split genes, and have the advantage of being capable to be expressed in bacteria, which do not have the machinery to process the eukaryotic mRNA.

There are far less number of cDNA clones in a bank than in a genomic library, which makes easier to look for a desired gene.

Screening of cDNA bank also provides fairly unambiguous results.

cDNA construction

Once mRNA is purified, oligo-dT (a short sequence of deoxy-thymidine nucleotides) is tagged as a complementary primer which binds to the poly-A tail providing a free 3'-OH end that can be extended by reverse transcriptase to create the complementary DNA strand. Now, the mRNA is removed by using a RNAse enzyme leaving a single stranded cDNA (sscDNA).



This sscDNA is converted into a double stranded DNA with the help of DNA polymerase. However, for DNA polymerase to synthesize a complementary strand a free 3'-OH end is needed. This is provided by the sscDNA itself by generating a hairpin loop at the 3' end by coiling on itself.

The polymerase extends the 3'-OH end and later the loop at 3' end is opened by the scissoring action of S1 nuclease.

Restriction endonucleases and DNA ligase are then used to clone the

sequences into bacterial plasmids.The cloned bacteria are then selected, commonly through the use of antibiotic selection. Once selected, stocks of the bacteria are created which can later be grown and sequenced to compile the cDNA library.

Genomic Library vs cDNA libraries

Genomic Library	cDNA libraries
It include all possible fragments of DNA from a given cell or organism.	cDNA library carries only expressed gene sequences.
It is larger	It is smaller
It represents the entire genome of an organism having both coding and non coding regions.	It represents only the expressed part of the genome and contain only coding sequences called ESTs
Expression of genes taken from genomic library is difficult in prokaryotic system like bacteria due to absence of splicing mechanism.	cDNA has only coding sequences therefore can be directly expressed in prokaryotic system.
Vectors used genomic library include plasmid, cosmid, lambda phage, BAC and YAC in order to accommodate large fragments	Vectors used cDNA library include plasmid, phagemids, lambda phage etc to accommodate small fragments as cDNA has no introns.
Uses of Genomic Library	Uses of cDNA Library

Lecture 9

Polymerase chain reaction (PCR)

PCR is a technique used in molecular biology.It provides us with a quick, accurate and efficient way of producing to amplify a single copy or a few copies of a segment of DNA to millions or even billions of copies to be able to an identify, manipulate DNA, detect infectious organisms, including the viruses that cause AIDS, hepatitis, tuberculosis, detect genetic variations, including mutations, in human genes and numerous other tasks. It is used in the early stages of processing DNA for sequencing, for detecting the presence or absence of a gene to help identify pathogens during infection, and when generating forensic DNA profiles from tiny samples of DNA. Briefly PCR replicates specific fragments between two primers, which are complementary to short segments of DNA.

How does PCR work?

The principles behind every PCR, whatever the sample of DNA is, are the same. Five core 'ingredients' are required to set up a PCR. These are:

1. The DNA template to be copied.
2. Pairs of DNA Primers, short stretches of DNA that initiate the PCR reaction, it is complementary to the flanking sequence of the target DNA fragment.
3. DNA nucleotide bases (deoxyribonucleoside triphosphates –dNTPs-). DNA bases are the building blocks of DNA and are needed to construct the new strand of DNA.
4. *Taq* polymerase enzyme to add in the new DNA bases.
5. Buffer to ensure the right conditions, it contain Mg^{+2} ions to providing a suitable chemical environment for optimum activity of *Taq*polymerase.

A single polymerase chain reaction can be divided into three stages:-

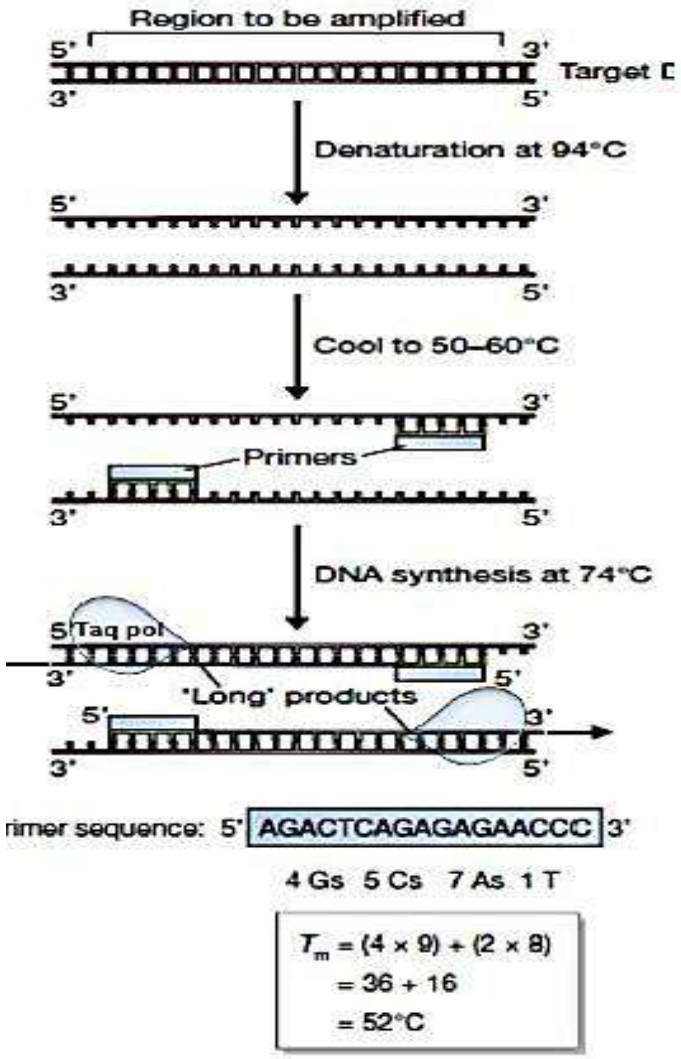
A) Denaturing (DNA strand separation) – when the double stranded template DNA is heated for long enough (up to 95 ^{0}C) .The high temperature causes the hydrogen bonds

between the bases in two strands of template DNA to break and the two strands to separated completely. This usually takes between 15-30 seconds.

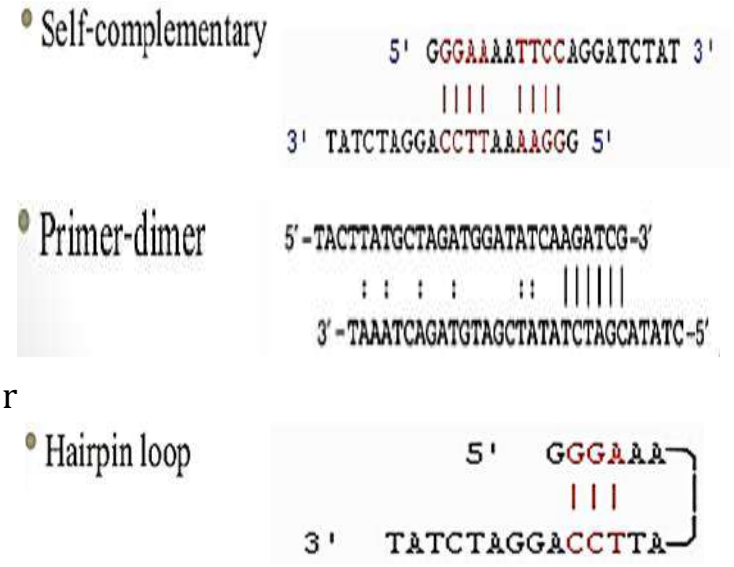
B) Annealing (DNA primer hybridization) – when the temperature is cooled to 50-65^{0}C to enable the DNA primers to hybridize with the 3' end of specific location on the single stranded template DNA by way of hydrogen bonding. The exact temperature dependson the melting temperature (Tm) of the primers. This step usually takes about 10-30 seconds.

Primer properties:

- Primers are single strands of DNA or RNA sequence that are around 17 to 28 bases in length and have a GC content of approximately 50%.
- The primers are designed to be complementary to short sections of DNA sequence on each end of the template single strands to be copied.
- **Avoid Cross Homology:** To improve specificity of the primers it is necessary to avoid regions of homology. Primers designed for a sequence must not amplify other genes in the mixture
- 3'end of primers should contain G or C or GC or CG that increase the efficiency of priming.
- Primers serve as the starting point for DNA synthesis.



- Primer self-complementarity (forming secondary structures such as hairpins when the 5' end binds to 3'end as containing palindromic ends) should be avoided in single primers. Also, the complementarity between the two primers should be avoided to prevent formation of primer dimers.



- A repeat is a di-nucleotide occurring many times consecutively and should be avoided because they can misprime. For example: ATATATAT. A maximum number of di-nucleotide repeats acceptable in an oligo is 4 di-nucleotides.

- The polymerase enzyme can only add DNA bases (dNTPs) to a double strand of DNA so the primer annealing provides that.

- Once the primer has bound, the polymerase enzyme can attach and start making the new complementary strand of DNA from the DNA bases (dNTPs) provided in the reaction buffer.

- The two separated strands of DNA are complementary and run in opposite directions (from 5’ end to 3’ end). As a result, there are two types of primers; a forward primer and a reverse primer.
- The annealing temperature of the primer pair is calculated from the equation Predictionof Tm. (above example)

$$T_m = (4 \times [G + C]) + (2 \times [A + T])^{\circ}C$$

C) Extending (DNA synthesis) –when the temperature is raised to 72^0C. This is the optimal temperature at which the heat-resistant DNA polymerase “Taq polymerase” begins to elongate and the new strand of DNA is made by this enzyme, which adds dNTPs to the single strand one-by-one in the 5’ to 3’ direction.

- *Taq* DNA polymerase is an enzyme extracted from thermophilic bacteria (Thermusaquaticus). This bacteria normally lives in hot springs so can tolerate temperatures above 80^0C. , which means DNA polymerase is very stable at high temperatures.

- DNA polymerases obtained from other organisms would not be able to tolerate these high temperatures; for example, human polymerases work ideally at 37˚C.

- *Taq* DNA polymerase needs divalent cations such as Mg^{+2} or Mn^{+2} at suitable concentration to function properly.

At low concentration of Mg^{+2}, the reaction fails because the polymerase is insufficiently active (no PCR product).

At high concentration of Mg^{+2}, the reaction loses specificity and multiple products are produced.

- The duration of this step depends on the length of DNA sequence being amplified but usually takes around one minute to copy 1,000 DNA bases (1Kb).

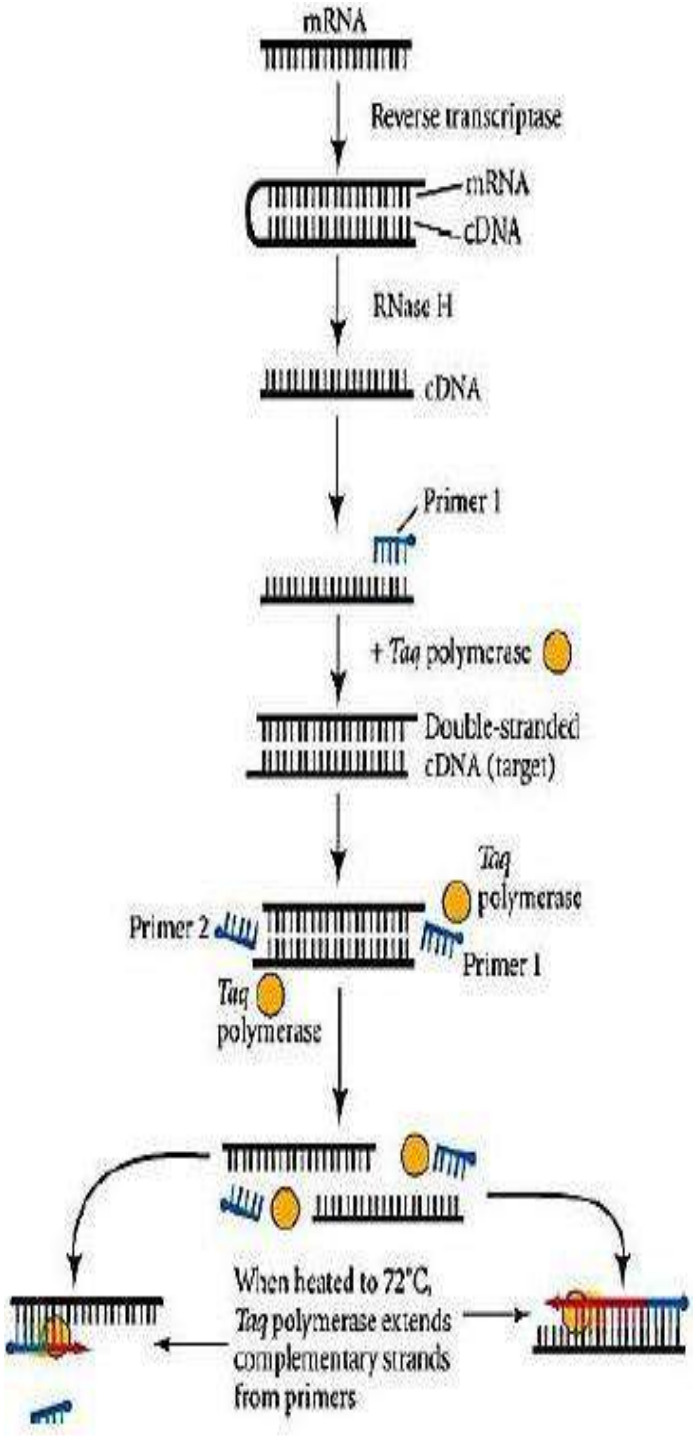
These three processes of thermal cycling are repeated 20-40 times to produce lots of copies of the DNA sequence of interest. The new fragments of DNA created during PCR serve as new templates for subsequent PCR cycles. The result is a huge number of copies of the specific DNA segment produced in a relatively short period of time.

C) Final elongation: this single step is performed at 70-74^{0}C (72^{0}C) for 5-15 minutes after the last PCR cycle to ensure that any remaining single stranded DNA is fully extended.

D) Final hold: this step at 4-15^{0}C for an indefinite time may be employed for short-term storage of the reaction.

Reverse transcription PCR (RT-PCR)

RT-PCR, also known as Reverse Transcriptase PCR, is a variation of the polymerase chain reaction that typically detection, amplification and measures RNA expression levels. The quality and purity of the RNA template is essential for the success of RT-PCR. The first step of RT-PCR is the synthesis of a DNA/RNA hybrid. In RT-PCR, complementary DNA (cDNA) is made by reverse transcribing of the RNA templates with the enzyme reverse transcriptase, Reverse transcriptase also has an RNase H function, which degrades the RNA portion of the hybrid. The single stranded DNA molecule is then completed by the DNA-dependent DNA polymerase activity of the reverse transcriptase into cDNA.



The possibility to revert RNA into cDNA by RT-PCR has many advantages. RNA is single-stranded and very unstable, which makes it difficult to work with. Most commonly, it serves as a first step in qPCR, which quantifies RNA transcripts in a biological sample. RT-PCR technique is used to qualitatively study gene expression, and can be combined with real time PCR (qPCR) to quantify

RNA levels. RT-PCR is used in research laboratories to study gene expression, for example in experiments to distinguish exons from introns, It can be adapted to identify mutations and polymorphisms and can be used clinically to diagnose genetic diseases and monitor drug therapy.

Lecture 10

qPCR and RT-qPCR

It is a quantitative method in contrast to conventional PCR, meaning that it enables the determination of exact amounts (relative or absolute) of amplified DNA in samples. Conversely, amplified DNA can only be detected after the amplification had been carried out (end-point detection) in conventional PCR.Apart from DNA, RNA can also be used as a template (e.g. in case of gene expression studies or detection of RNA viruses). In this case the RNA needs to be reverse transcribed into DNA (also termed complementary DNA or cDNA) before it is amplified with real-time PCR. As in standard PCR, DNA is amplified by 3 repeating steps: denaturation, annealing and elongation. However, in qPCR, fluorescent labeling enables the collection of data as PCR progresses

Two types of detection chemistry are used for real-time PCR analysis.

1- Fluorescent dyes, SYBR Green dye is used most commonly. This detection method is suitable when a single amplicon is being studied, as the dye will intercalate into any double-stranded DNA generated. During each cycle, the fluorescence is measured. The fluorescence signal increases proportionally to the amount of replicated DNA and hence the DNA is quantified in "real time".

- The advantages of SYBR Green dye are that it is inexpensive, easy to use, and sensitive. There is no need to design a probe for any particular target being analyzed. However, detection by SYBR Green requires extensive optimization.

- The disadvantages is that SYBR Green will bind to any double-stranded DNA in the reaction, including primer-dimers and other non-specific reaction products, which results in an overestimation of the target concentration.

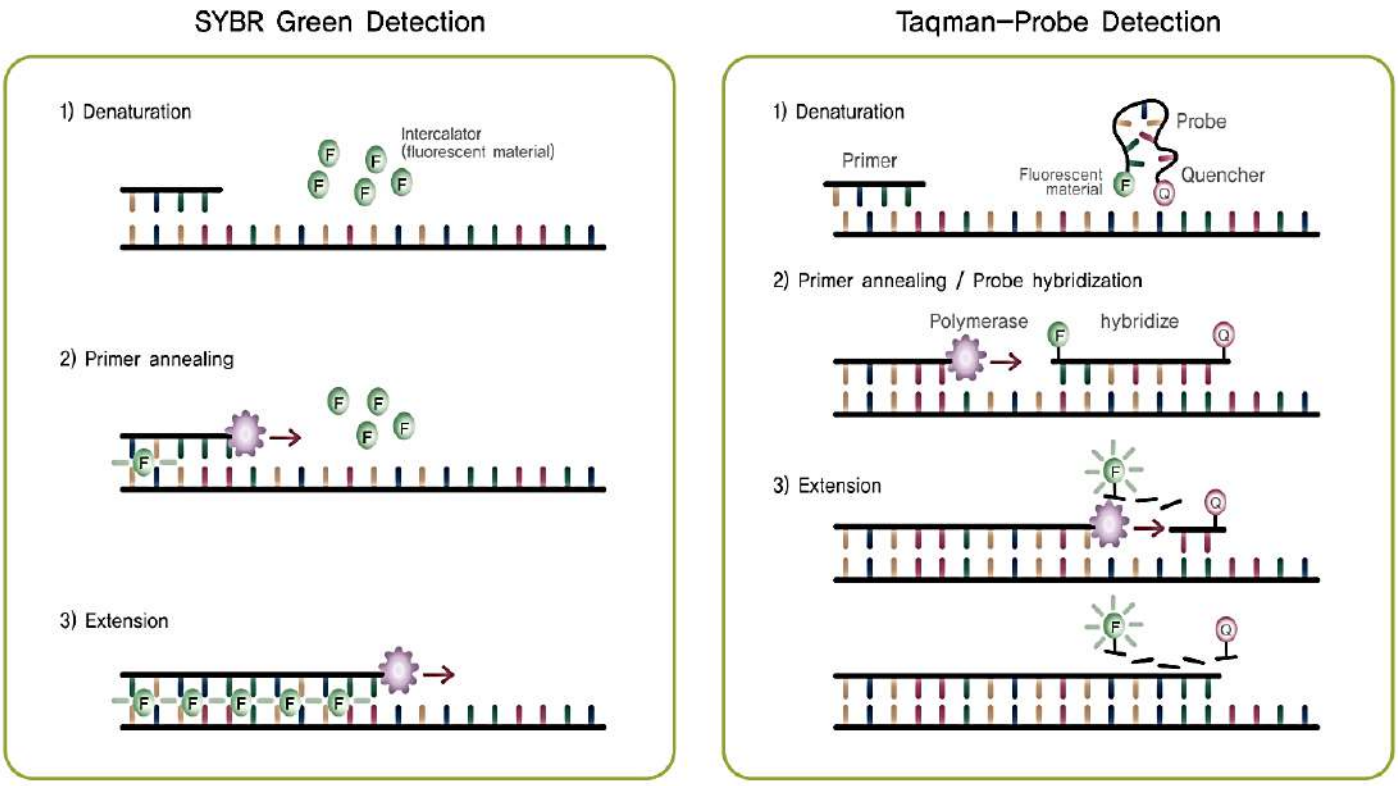
2- TaqMan probes are oligonucleotides that have a fluorescent reporter dye attached to the 5' end and a quencher moiety coupled to the 3' end.

These probes are designed (used in addition to primers) to hybridize to an internal region of a PCR product. In the unhybridized state, the proximity of the Fluor and the quench molecules prevents

the detection of fluorescent signal from the probe. During PCR, when the polymerase replicates a template on which a TaqMan probe is bound, the 5'- nuclease activity of the polymerase cleaves the probe. Thus, fluorescence increases in each cycle, proportional to the amount of probe cleavage.

Well-designed TaqMan probes require very little optimization. In addition, they can be used for multiplex assays by designing each probe with a spectrally unique fluor/quench pair. However, TaqMan probes can be expensive to synthesize, with a separate probe needed for each mRNA target being analyzed.

Note:Because probe-based qPCR is more specific than dye-based qPCR, it is often the technology used in qPCR diagnostic assays.



Advantages of Real Time PCR

Real Time PCR has many advantages over the normal PCR:

- It gives a look in to the reaction that is help to decide which reactions have worked well and which have failed.
- The efficiency of the reaction can be precisely calculated.
- There is no need to run the PCR product out on a gel after the reaction as the melt curve analysis serve the purpose.
- The real-time PCR data can be used to perform truly quantitative analysis of gene expression. In comparison, old fashioned PCR was only ever semi-quantitative at best.
- Faster than normal PCR.
- Less complexity at the quantification of sample.etc.

Thus, unlike the ordinary preparative PCR, RT-PCR allows the success of multiple PCR reaction to be determined automatically after only a few cycles, without separate analysis of each reaction, and avoids the problem of "false negatives".

In general the Applications of Real-Time PCR are include:

- Gene Expression (mRNA) Analysis
- microRNA and Non-Coding RNA analysis
- Genetic Variation
- Mutation Detection
- SNP Analysis
- Genotyping/Allelic Discrimination

RAPD and RFLP

Genetic markers are used in Molecular Biology to identify genetic variations between individuals and species such as:

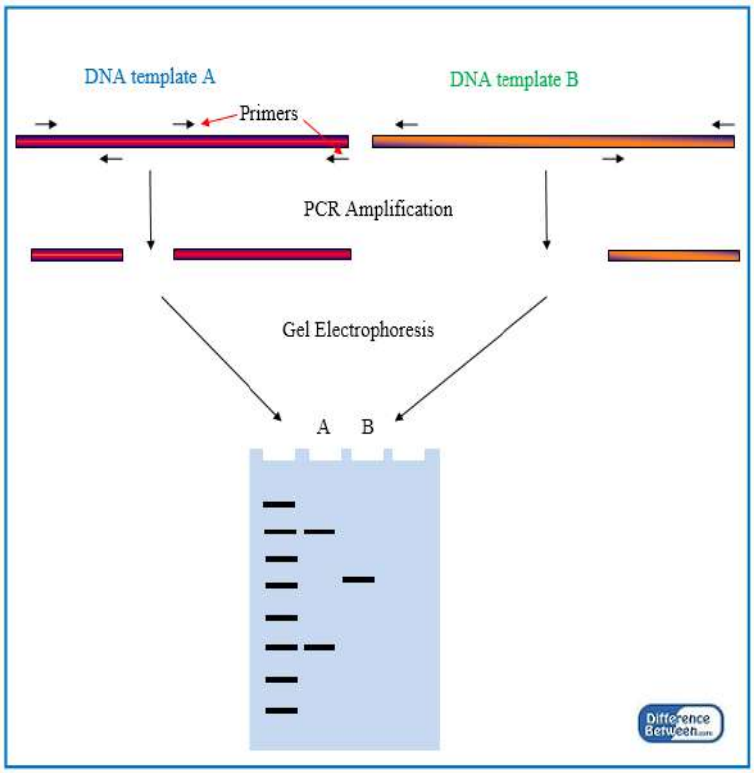
1- **Random Amplified Polymorphic DNA (RAPD):** It is a quick and easy technique. RAPD can be defined as a method which results in polymorphic DNA sequences as a result of random

amplification of multiple locations of the target DNA template. RAPD uses short oligonucleotide primers with arbitrary sequences for the PCR amplification. Primers are artificially synthesized without the prior sequence knowledge. Hence, it is considered as an easy and useful technique.

The following major steps are involved in RAPD.

- Extraction of target DNA
- Amplification of the multiple locations of the target DNA using randomly chosen primers.
- Gel electrophoresis of the amplified PCR products.
- Staining with ethidium bromide and identification of the polymorphism

As a result of the variation in the primer annealing, different fragments with different lengths are generated during the amplification. Hence, banding patterns on the gels are different among individuals and species. Thus, RAPD enables the detection of genetic variation among organisms in identification and differentiation.



RAPD is applied in various studies of molecular biology such as identification of the genetic difference between closely related species, gene mapping, DNA fingerprinting, identification of inherited diseases, etc

2- **Restriction Fragment Length Polymorphism (RFLP):** It is the first genetic marker developed for DNA fingerprintingand identification of genetic variation in homologous DNA sequences. When DNA samples are digested with specific restriction endonucleases, it yields

different DNA profiles which are unique to each individual and analysis of the fragment length polymorphism via gel electrophoresis and blotting. Blotting patterns are unique to each organism and characterize the specific genotypes.

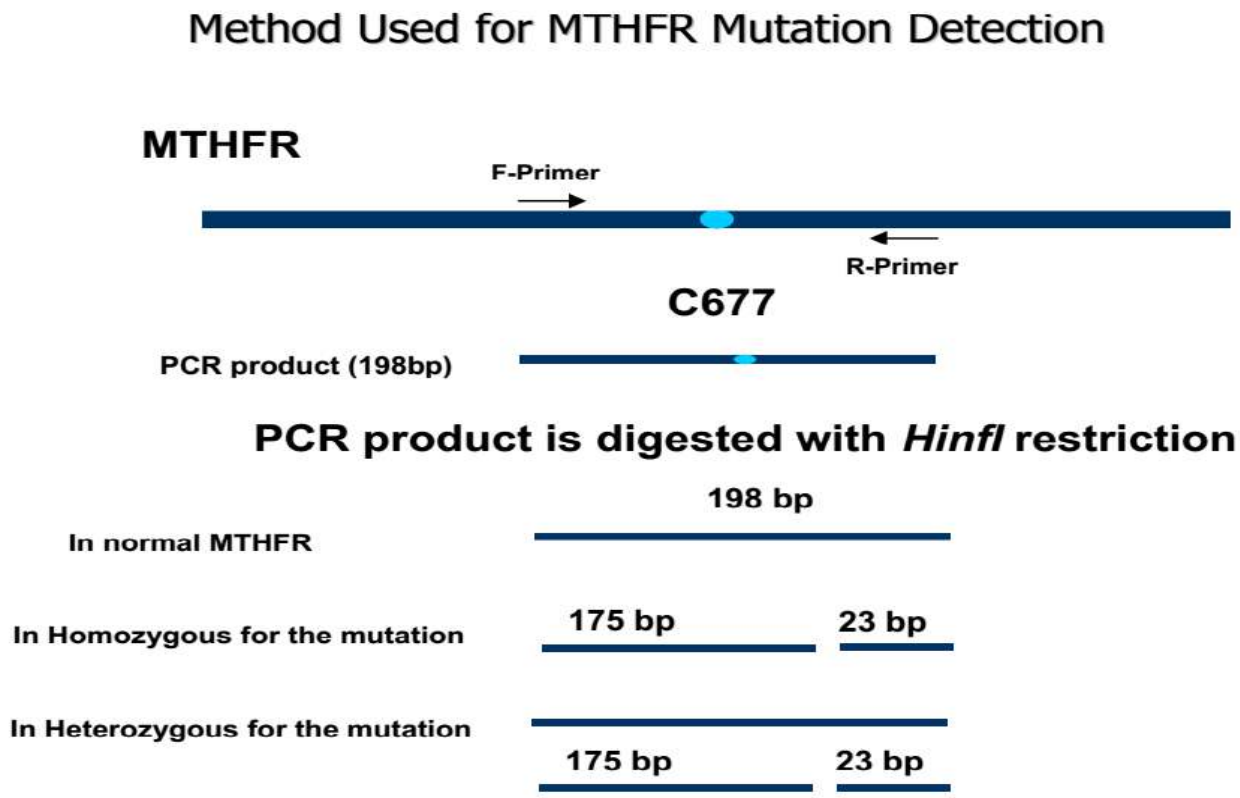
Following steps are involved with RFLP.

1. Isolation of sufficient amount of DNA from samples
2. Fragmentation of the DNA samples with specific restriction endonucleases into short sequence
3. Separation of the resulted fragments with different lengths by agarose gel electrophoresis.
4. Transfer of the gel profile into a membrane by Southern blotting
5. Hybridization of the membrane with labeled probes and analysis of the fragment length polymorphism in each profile

RFLP has various applications such as diagnosis of inheritance diseases, genome mapping, criminal identification in forensic studies, paternity testing, etc.

Example of use of RFLP in the detection of mutation in disease state:

Mutation in the Methyl tetrahydrofolate reductase enzyme could lead to increased level of homocystein in blood and leads to increased risk of thrombosis in these individuals carrying the mutation

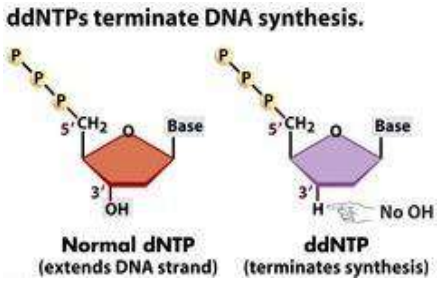


Lecture 11

DNA sequencing

Is the process to tell the order of the bases, or nucleotides, that form the inside of the double-helix molecule by making many copies of a target DNA region. Its ingredients are similar to those needed PCR technique .They include:

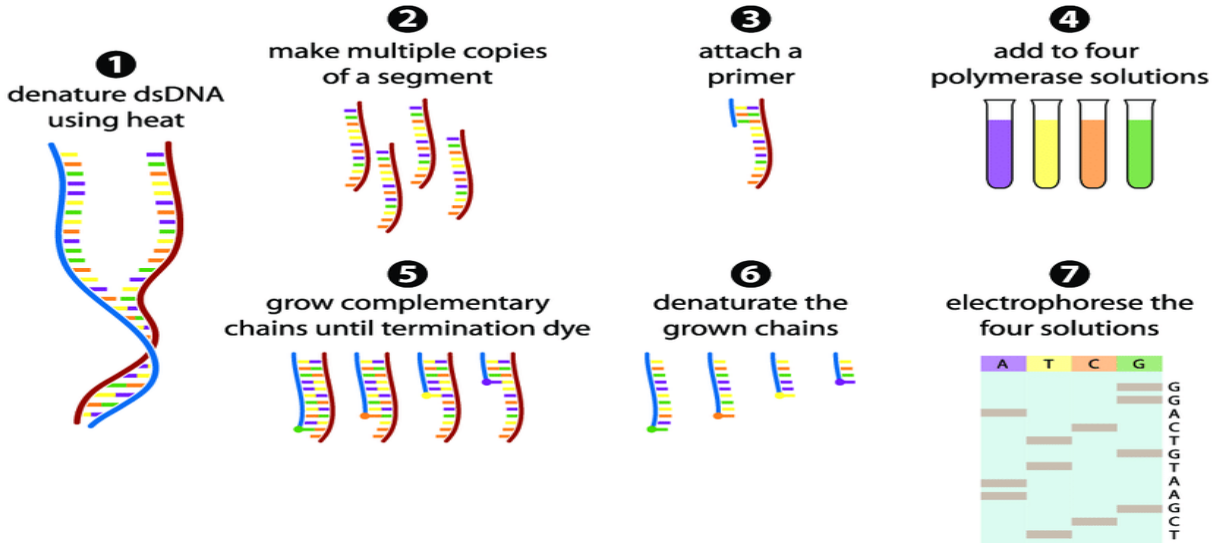
- The template DNA to be sequenced
- DNA polymerase.
- Primers.
- radioactively or fluorescentlydNTPs (dATP, dTTP, dCTP, dGTP), which are deoxy or extends
- dideoxynucleotide triphosphates (ddNTPs) as DNA chain terminators , It lack a 3'-OH end (that terminate DNA strand elongation).



Sequencing reaction

The mixture is first heated to denature the template DNA (separate the strands), then cooled so that the primer can bind to the single-stranded template. Once the primer has bound, the temperature is raised again, allowing DNA polymerase to synthesize new DNA starting from the primer. DNA polymerase will continue adding nucleotides to the chain until it happens to add a ddNTPs instead of a normal one. At that point, no further nucleotides can be added, so the strand will end with the ddNTPs.

Following synthesis, the products of the A, G, C, and T reactions are individually loaded into four lanes of a single gel and separated using gel electrophoresis, a method that separates DNA fragments by their sizes. The DNA bands are then visualized by autoradiography or UV light, and the DNA sequence can be directly read off the X-ray film or gel image from the bottom to the topof the gel, including bands in all four lanes.

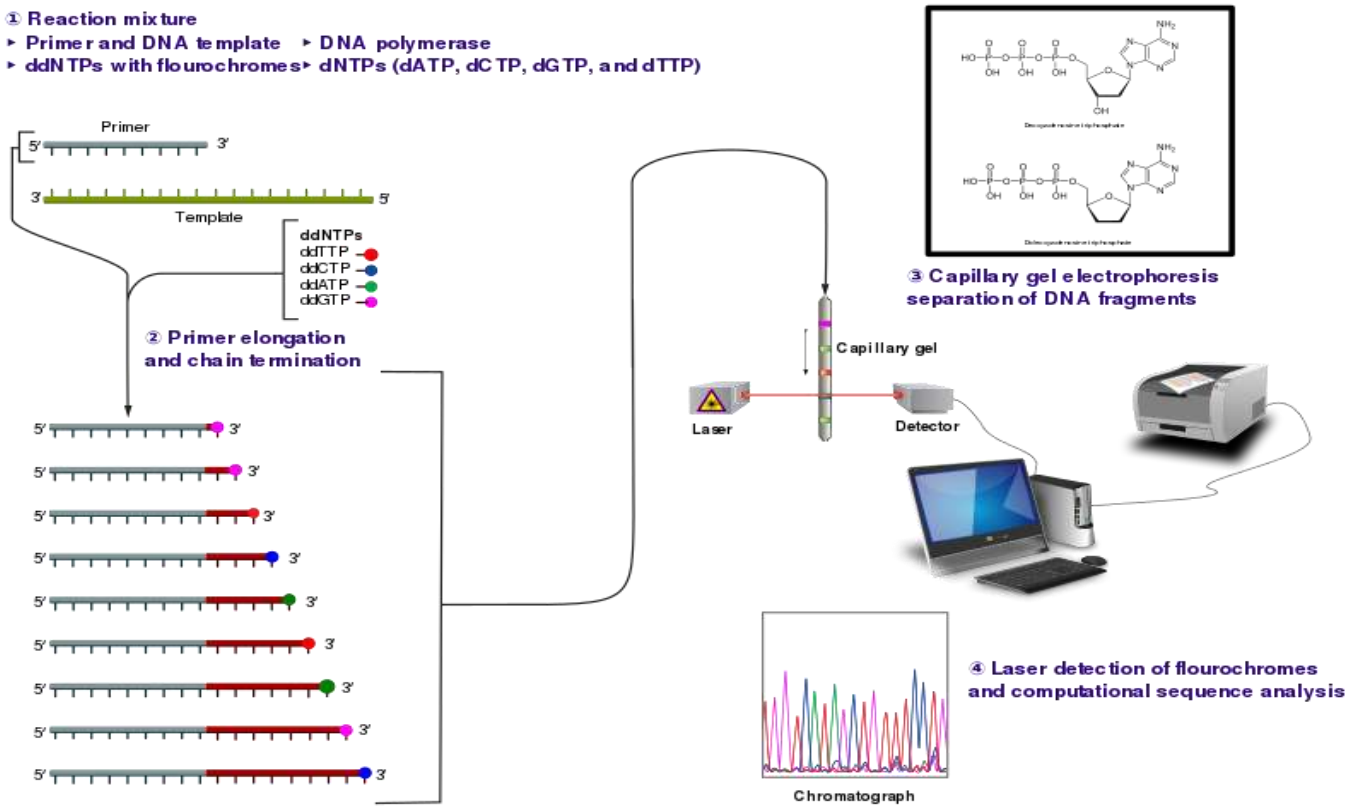


⇒Sanger sequencing was used to determine the sequences of many relatively small fragments of human DNA. (These fragments weren't necessarily 900 bp or less, but each fragment using multiple rounds of Sanger sequencing). The fragments were aligned based on overlapping portions to assemble the sequences of larger regions of DNA and, eventually, entire chromosomes.

✓ The common challenges of DNA sequencing include poor quality in the first 15-40 bases of the sequence and deteriorating quality of sequencing traces after 700-900 bases.

Cycle sequencing

In cycle sequencing, the reaction mixture containing template DNA, primer, DNA polymerase, dNTPs, and all four fluorescently labelled ddNTPs is subjected to about 30 cycles of annealing, extending, and melting (Figure 3). Although incorporation of a dNTP will continue to grow the chain, a ddNTP will block the extension, resulting in DNA fragments of varying lengths. After cycle-sequencing reaction, PCR products are purified manually and size separated by automated capillary gel electrophoresis, coupled with a laser detector located at the end of the capillary. The laser detector is capable of detecting all four fluorescent dyes as the migrating fragments pass through the analysis window. Because each ddNTP has a different colour, the sequence can be determined as chromatogram by correlating the colour of a band on the gel with its specific ddNTP, hence the order in which they ran on the gel.



Next-Generation DNA Sequencing (NGS)

Next-generation sequencing (NGS), also known as high-throughput sequencing, is used to describe a number of different modern sequencing technologies including:

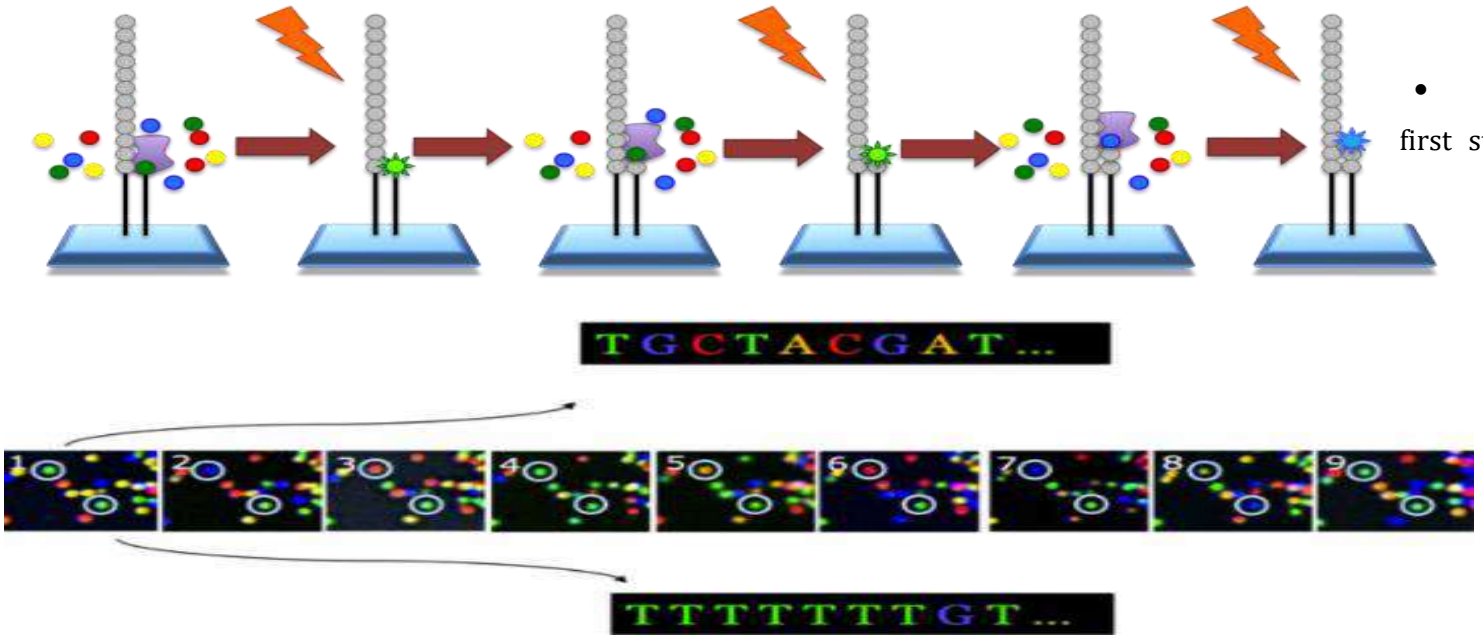
1- Illumina (Solexa) sequencing 2- Roche 454 sequencing 3- Ion torrent: Proton / PGM sequencing

Illumina sequencing

Vast numbers of short reads are sequenced in a single stroke. To do this, firstly the input sample must be cleaved into short sections. In Illumina sequencing, 100-250bp reads are used. Somewhat longer fragments are ligated to generic adaptors and annealed to a slide using the adaptors. PCR is carried out to amplify each read, creating a spot with many copies of the same read. They are then separated into single strands to be sequenced. The slide is flooded with nucleotides (fluorescently labelled, with the colour corresponding to the base)and DNA polymerase.These nucleotides also have a terminator, so that only one the base is added at a time.

The slide is then prepared for the next cycle. The terminators are removed, allowing the next base to be added,and the fluorescent signal is removed, preventing the signal from contaminating the next image.

Computers are then used to detect the base at each site in each image and these are used to construct a sequence. The process is repeated, adding one nucleotide at a time and imaging in between.All of the sequence reads will be the same length, as the read length depends on the number of cycles carried out.



- The first step of the sequencing technology is to randomly fragment the genomic DNA due the fact that the sequence reads with more than 250 bp for each read cannot be sequenced properly because terminator and fluorescent signal cannot be removed leading to contaminate the next image.
- Secondly the adaptor (20 - 50 bps) is linked to each fragment and the adaptor help to link the DNA fragment on slides. There are two adaptors are used forward and reverse adaptor to sequence each fragment from both direction (The forward and reverse direction). Each fragment will be amplified using bridge amplification to increase the number of reads for each fragment and this is called depth of the coverage. Depth of coverage represents the number of read for each fragment and if the depth of the coverage increases the accuracy of the sequencing increase as well.
- Finally the reads will be aligned to a reference genome for both the forward and reverse read..

The advantages of NGS over classical Sanger sequencing are:Speed, Cost, Sample size, Accuracy.

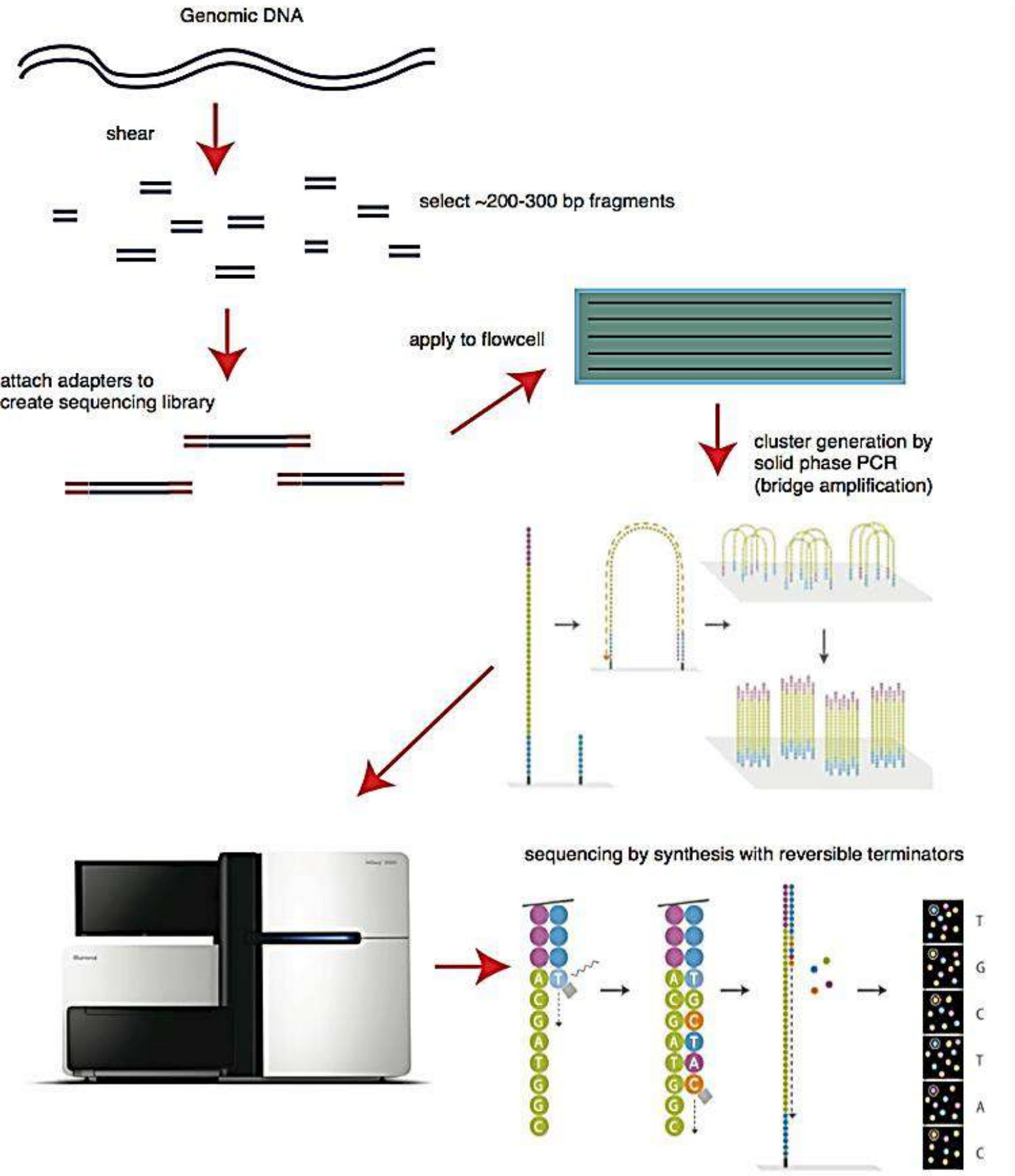
Sanger sequencing,a large amount of template DNA is needed for each read.

Several strands of template DNA are needed for each base being sequenced (i.e. for a 100bp sequence you'd need many hundreds of copies, for a 1000bp the sequence you'd need many thousands of copies), as a strand that terminates on each base is needed to construct a full sequence. But for NGS, a sequence can be obtained from a single strand.

NGS is quicker than Sanger sequencing in two ways.

1- The chemical reaction may be combined with the signal detection in some versions of NGS, whereas in Sanger sequencing these are two separate processes.
2- Only one read (maximum ~1kb) can be taken at a time in Sanger sequencing, whereas NGS allowing 300Gb of DNA to be read on a single run on a single chip.
3- It is possible to do more repeats than with Sanger sequencing.

Repeats are intrinsic to NGS, as each read is amplified before sequencing, and because it relies on many short overlapping reads, so each section of DNA or RNA is sequenced multiple times. Also,More repeats mean greater coverage, which leads to a more accurate and reliable sequence, even if individual reads are less accurate for NGS. Sanger sequencing can be used to give much longer sequence reads. However, the parallel nature of NGS means that longer reads can be constructed from many contiguous short reads.


Genomic DNA
shear
select ~200-300 bp fragments
attach adapters to
create sequencing library
apply to flowcell
cluster generation by
solid phase PCR
(bridge amplification)
sequencing by synthesis with reversible terminators
T
G
C
T
A
C

Lecture 12

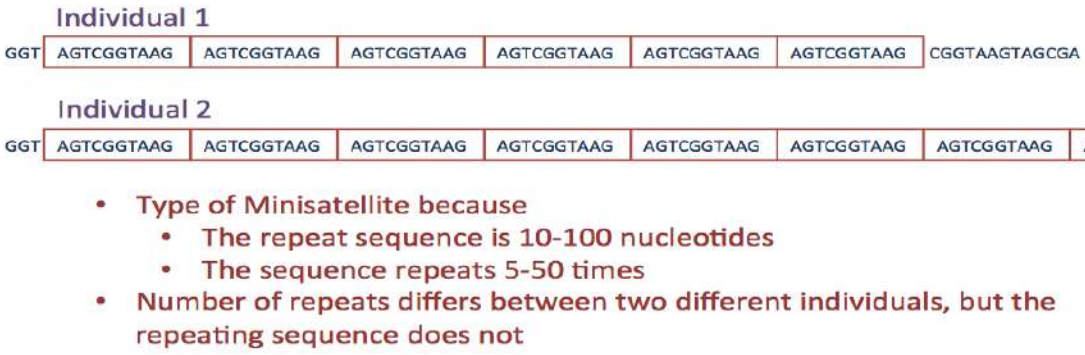
Mapping Genomes:

It is the process of finding the locations of genes on each chromosome by using genetic markers. A genetic marker is a gene or sequence on a chromosome that co-segregates (shows genetic linkage) with a specific trait.

A good genetic marker is a region on the chromosome that shows variability or polymorphism in the population. Some genetic markers used in generating genetic maps are:

1- Restriction fragment length polymorphisms (RFLP)
2- Variable number of tandem repeats (VNTRs): repeated sets of nucleotides present in the non-coding regions of DNA called, microsatellite polymorphisms.
3- Single nucleotide polymorphisms (SNPs).

There are two distinctive types of "Maps" used in the field of genome mapping: genetic maps and physical maps. While both maps are a collection of genetic markers and gene loci.



SNP

A single nucleotide polymorphism, is a variation at a single position in a DNA sequence among individuals in coding and non-coding region of DNA.

If a SNP occurs within coding regions, then the gene is described as having more than one allele. In these cases, SNPs may lead to variations in the amino acid sequence.

Most SNPs have no effect on health or development. Importance of SNPs are:

1- Some of these genetic differences, have proven to be very important in the study of human health.
2- Help predict an individual's response to certain drugs, susceptibility to environmental factors such as toxins, and risk of developing particular diseases.
3- SNPs can also be used to track the inheritance of disease genes within families. Future studies will work to identify SNPs associated with complex diseases such as heart disease, diabetes, and cancer.
4- SNPs are used for identification and forensics
5- SNPs are used for mapping

6- SNPs are used to predict specific genetic traits

Genetic Maps

The study of genetic maps are based on the genetic linkage information,a procedure that analyses the recombination frequency between genes to determine if they are linked or show independent assortment.

The generation of genetic maps requires markers(More ***advanced markers including those based on non-coding DNA),*** are now used to compare the genomes of individuals in a population. Although individuals of a given species are genetically similar, they are not identical; every individual has a unique set of traits.These minor differences in the genome between individuals in a population are useful for the purposes of genetic mapping.

In general, genome mapping can be used in a variety of other applications:

- ✓ Having a complete map of the genome makes it easier for study individual genes, such as Human genome maps help to identify human disease-causing genes like cancer, heart disease, and cystic fibrosis.
- ✓ Genome mapping can be used in a variety of other applications, such as using live microbes to clean up pollutants or even prevent pollution.
- ✓ Plant genome mapping may lead to producing higher crop yields or developing plants that better adapt to climate change.

Because genetic maps rely completely on the natural process of genetic recombination, So; mapping is affected by natural increases or decreases in the level of genetic recombination in any given area of the genome. Some parts of the genome are hotspots (increased recombination), whereas others do not show a propensity for recombination. For this reason, it is important to look at mapping information developed by multiple methods.

Physical maps

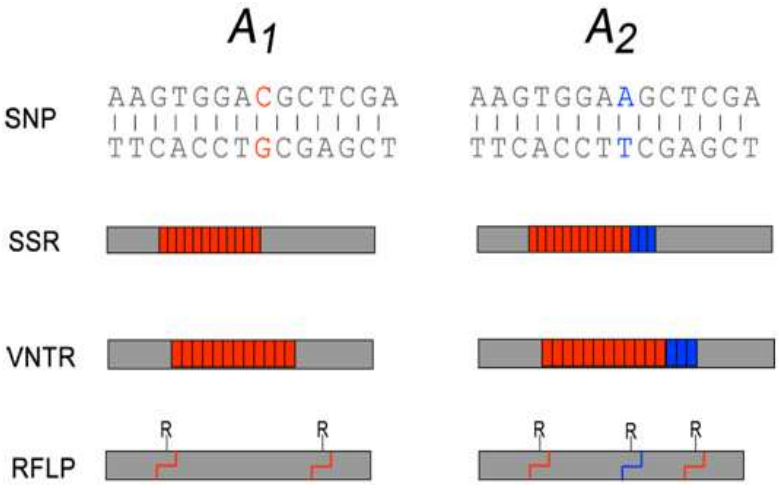
Physical maps give an idea of the actual distance of the genetic marker and the distance of the base pairs on a gene. Identifying the location of the gene helps us understand the nature of disease, whether the disease is genetically inherited or is caused by just random mutation on the gene.Physical maps are constructed using three methods:

1- Cytogenetic mapping: It uses information obtained by microscopic analysis of stained sections of the chromosome. It is possible to determine the approximate distance between genetic markers using cytogenetic mapping, but not the exact distance (number of base pairs).

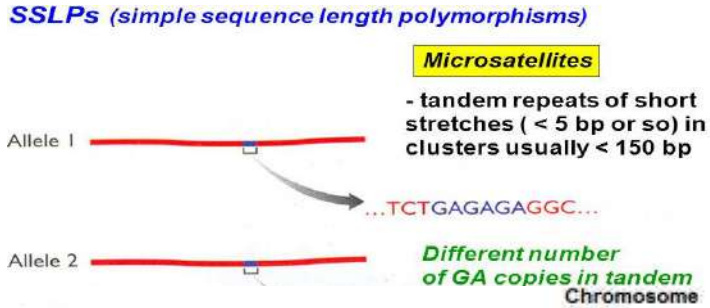
2- Radiation hybrid mapping: It uses radiation, such as x-rays, to break the DNA into fragments. The amount of radiation can be adjusted to create smaller or larger fragments. This technique overcomes the limitation of genetic mapping and is not affected by increased or decreased recombination frequency.

3- Sequence mapping: It resulted from DNA sequencing technology. The genomic libraries and cDNA libraries has sped up the process of physical mapping.

In general a genetic site used to generate a physical map with sequencing technology (a sequence-tagged site, or STS) is aunique sequence in the genome (100-500bp in length) that is easily recognized, occurs only once in the chromosomal or genome. An expressed sequence tag (EST) and a single sequence length polymorphism (SSLP) are common STSs.



An EST is a short STS that is identified with cDNA libraries, while SSLPs are obtained from known genetic markers and provide a link between genetic maps and physical maps.



Integration of Genetic and Physical Maps

Genetic maps provide the outline and physical maps provide the details. It is easy to understand why both types of genome mapping techniques are important to show the big picture. Information obtained from each technique is used in combination to study the genome. Genomic mapping is being used with different model research organisms. Genome mapping is an-ongoing process; as better techniques are developed, more advances are expected.