

VU LEARNING HUB

vulearninghub.xyz

HIGHLIGHTED HANDOUT

BT301

Introduction to Biotechnology

BT301 by VU Learning HUB

Prepared for VU students • Free study resources

<https://vulearninghub.xyz>

Virtual University Of Pakistan

Subject Bt301

Introduction of Biotechnology

Email Bt301@vu.edu.pk

Create By

VU Bio Mates Team

<https://www.facebook.com/groups/vubiomates/>

01 Introduction to Biotechnology

Biotechnology is the use of microbes, animal/plant cells and their products to synthesize, break down or transform materials. Primarily it includes the use of recombinant DNA technology and genetic engineering techniques to improve upon the quality of processes.

Traditional biotechnology refers to the conventional techniques that have been used for many centuries to produce beer, wine, cheese etc.



Modern Biotechnology embraces all methods of genetic modification by recombinant DNA & cell fusion techniques together with the modern developments of traditional biotechnological processes.



White Biotechnology

Development of processes and microorganisms for Industrial processes.

Key concept

Scale of operation Low volume high price

White Biotechnology

Example

Enzyme production image1



Red Biotechnology

It is concerned with the discovery and development of innovative drugs and treatments. Image2

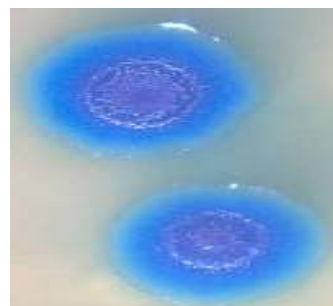
Green Biotechnology

It is concerned with the modification of the genetic composition of plants to enhance existing traits or add new ones.

Example

Bt corn from Syngenta

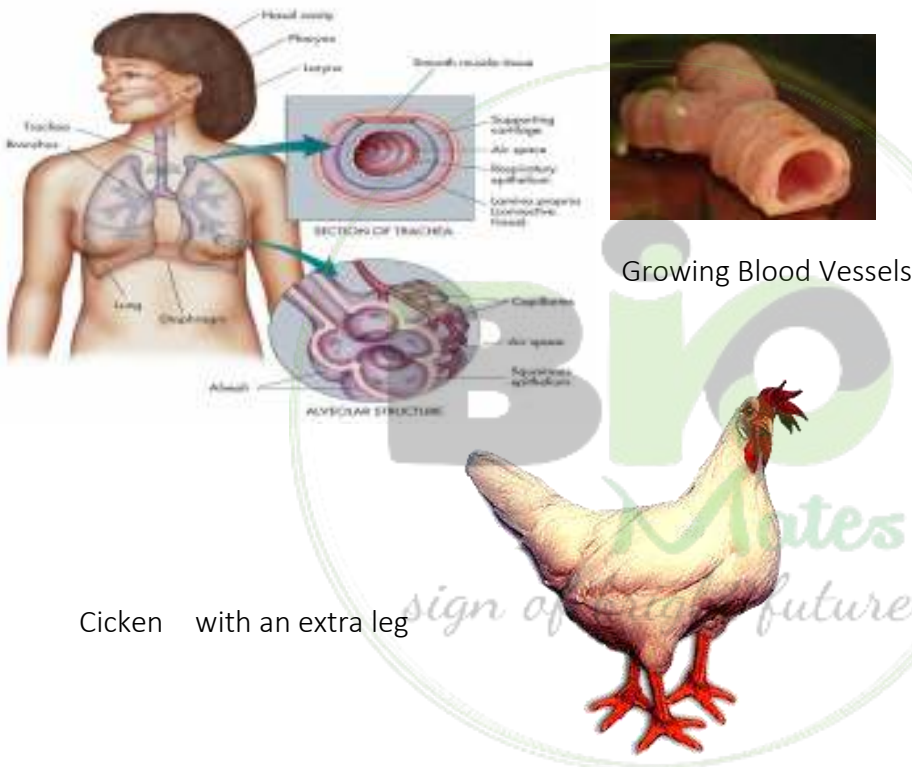
Bt cotton from Monsanto



Growing Teeth



Claudia's Trachea



Chicken with an extra leg



Growing Meat



03 Roots of Biotechnology

6000 BC

Sumerians & Babylonians used anaerobic respiration of yeast to make beer.

4000 BC

Egyptians used aerobic respiration of yeast to leaven or rise bread and used mold to flavour cheese.

1350 BC

1350 BC

Pregnancy test



Barley grows – male child
Wheat grows – female child

400 BC

Use of salts to preserve food specially meat.



1676 AD

Antonie Van Leeuwenhoek
Observation of microscopic single celled organisms



1828 AD

Synthesis of Urea
Fredrick Wohler
Pioneer in inorganic chemistry for synthesizing urea. This discovery became refutation of vitalism, the hypothesis that living things are alive because of some special "vital force" 



1834 - 1839 AD

Cell Theory
Theodore Schwann
"All living things are composed of cells and cell products"



1857 AD

Rudolf Virchow
Omnis cellula e cellula
"Every cell arise from another cell"



03 Discovery of DNA

1869

Friedrich Miescher
discovered nuclein

1878

Albrecht Kossel

Isolated non-protein component of nuclein & discovered nucleic acids & isolated 5 bases: A, T, G, C & U

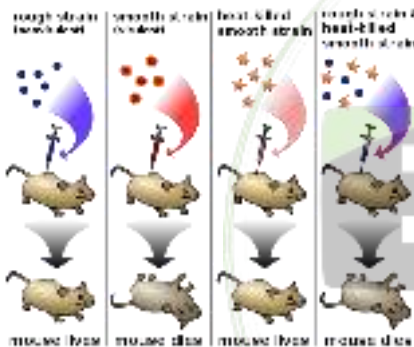
1927

Nikolai Konstantinovich Koltsov

Inherited traits would be inherited via a "giant hereditary molecule" which would be made up of "two mirror strands that would replicate in a semi-conservative fashion using each strand as a template"



1928 Frederick Griffith Discovered transformation, proof that DNA contains genetic information, different strains of streptococcus pneumoniae, Type 3S and Type 2R



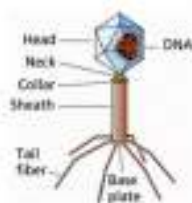
1944

Avery, MacLeod & McCarty Experiment



1952

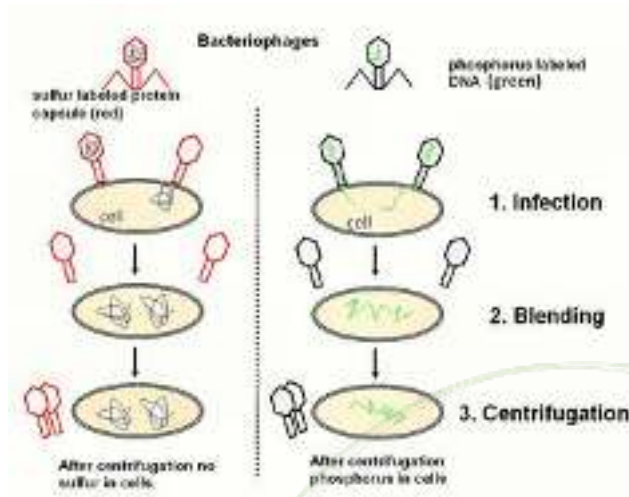
Hershey-Chase



Experiment

1952

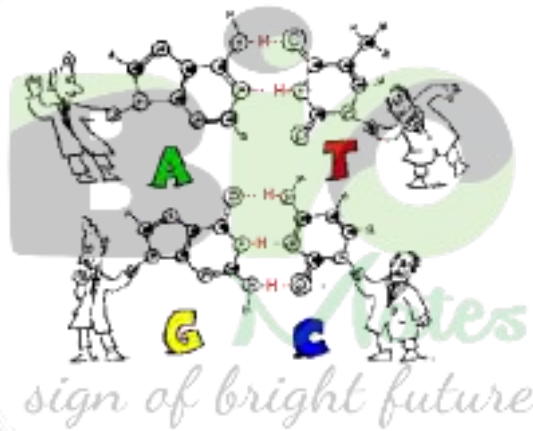
Hershey-Chase Experiment



Watson & Crick 1953



Watson & Crick 1953



06 Dawn of Biotechnology



Eye colour



curly hair



tongue roller

1973

Boyer brought in restriction enzymes

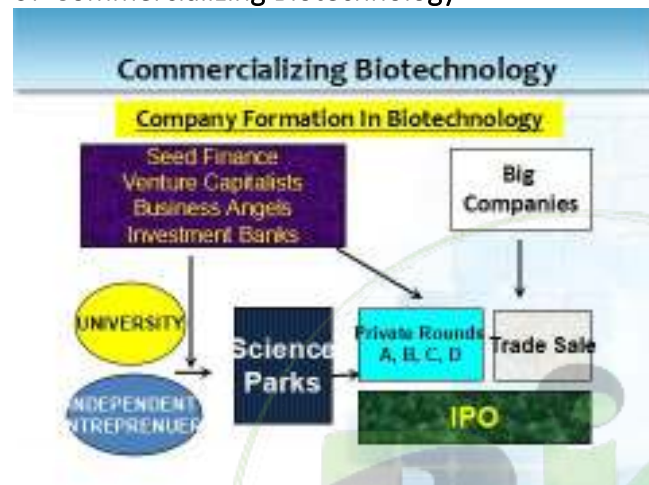
Cohen contributed with plasmid (pSC101)

African Clawed Frog gene into a bacterium

Negative Perception

“Genetic Engineering can give rise to something which can crawl out of the laboratory such as Frankenstein”

07 Commercializing Biotechnology



Commercializing Biotechnology

First generation of Biotech Companies

Company	Date of IPO	Amount Raised
Genentech	10/80	\$35M
Cetus	03/81	\$107M
Genetic Systems	04/81	\$6M
Pfizer Immunochem	05/81	\$1.8M
Genome Therapeutics	05/82	\$12.9M
Centocor	12/82	\$21M
Scios	01/83	\$12M

Commercializing Biotechnology

First generation of Biotech Companies

Company	Date of IPO	Amount Raised
Biotechnology General	09/83	\$8.9M
Immunex	03/83	\$16.5M
Amgen	06/83	\$42.3M
Biogen	06/83	\$57.5M
Chiron	08/83	\$17M
Immunomedics	11/83	\$2.5M
Regeneron	04/86	\$17.5M

08 Commercializing Biotechnology

Roche acquired Genentech for \$46.8 billion
Current employees = 12,895



19 Breakthrough Biopharmaceutical Products



Revenue \$20 billion

Total Assets \$69 billion

net Income \$ 5 billion

Current employees = 17,900

11 Breakthrough Biopharmaceut



Revenue \$9.7 billion

Total Assets \$15 billion

Net Income \$ 3 billion

Current employees = 7,550

10 Breakthrough Biopharmaceutical Products



Net Income \$ 421 million

Current employees = 12,000

16 innovative Biopharmaceutical Products



Gaucher's Disease

Genetic disease in which fatty substances accumulate in cells and certain organs

Molecular Reason:

Hereditary deficiency of the enzyme glucocerebrosidase

Symptoms:

Bruising, Fatigue, Anemia, low blood platelets, enlargement of the liver and spleen

Prior to this, they had to collect 22,000 placentas to treat Gaucher's patients

11 Nucleoside & Nucleotid

A molecule containing all these three components is called a nucleotide.

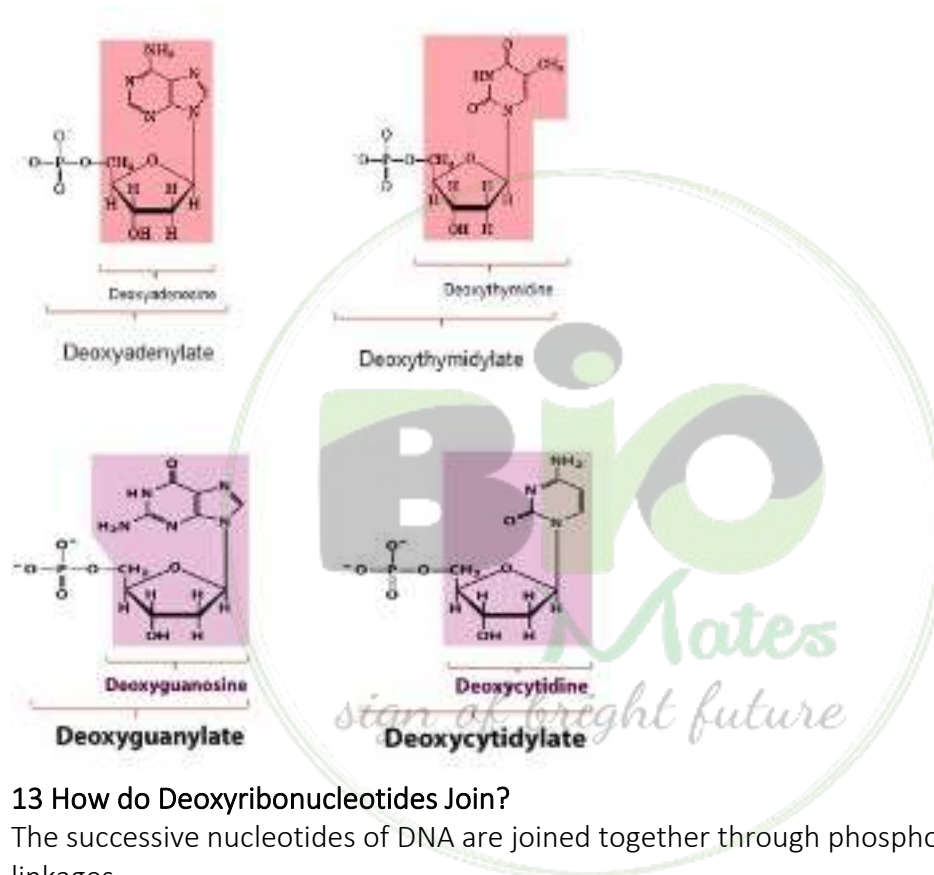
While a molecule without the phosphate group is called a nucleoside



Nucleotide = Nucleoside + Phosphoric acid

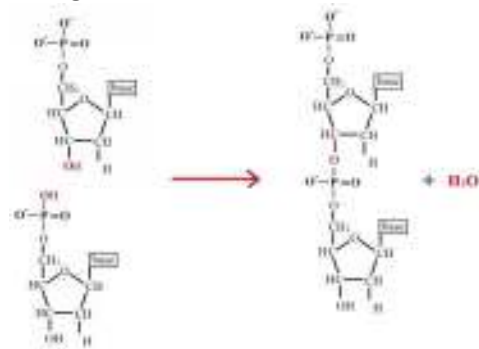
Nucleoside = Nucleotide – Phosphoric acid

12 Types of Deoxyribonucleotides



13 How do Deoxyribonucleotides Join?

The successive nucleotides of DNA are joined together through phospho-diester linkages.



14 Structure of DNA

Work of Chargaff (Late 1940s)

The discovery of the structure of DNA is one of the greatest events in the history of science.

Erwin Chargaff and his colleagues provided a most important clue to the structure of DNA.

The work of Chargaff led him to following conclusions, also called “Chargaff Rules”:-

1. Base composition of DNA varies from one species to another.

2. The DNA isolated from different tissues of the same species have the same base composition.
3. The base composition of DNA in a given species does not change with an organism's age, nutritional state, or changing environment.
4. In DNA, the number of adenosine residues is equal to the number of thymidine (A=T) and the number of guanosine residues is equal to the number of cytidine (G=C).
5. It means that the sum of the purine residues equals the sum of the pyrimidine residues (AG=TC).

15 Work of Franklin & Wilkins (1950s)

Rosalind Franklin and Maurice Wilkins performed the x-ray diffraction analysis of DNA fibers.

They showed that DNA produces a characteristic x-ray diffraction pattern.

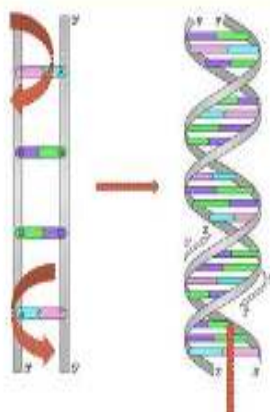


From this pattern, they made two important findings:-

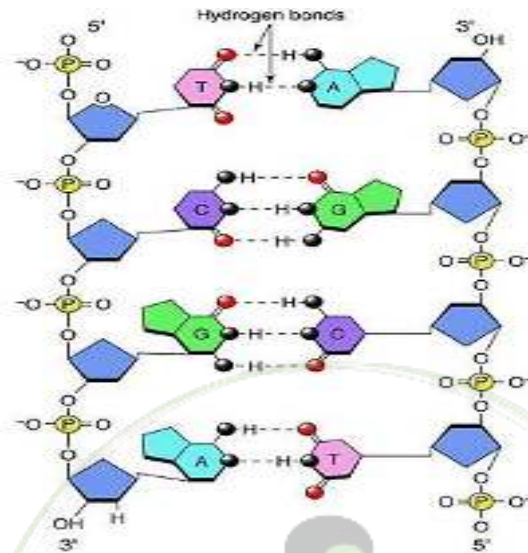
1. DNA molecules are helical.
2. The helices have two periodicities along their long axis, a primary one of 3.4 Å and a secondary one of 34 Å.

16 Work of Watson & Crick

James Watson and Francis Crick postulated a three dimensional model of DNA structure in 1953. The major features of this model are as follow:-



1. DNA consists of two helical polynucleotide strands which are wound around the same axis to form a right handed double helix.
2. The hydrophilic backbones of alternating deoxyribose and phosphate groups are on the outside of the double helix, facing the surrounding water.
3. The nitrogenous bases of both strands are stacked inside the double helix lying perpendicular to the long axis of the helix.
4. The pairing and coiling of the two strands create a major groove and minor groove on the surface of the helix.



17 Chemical composition of RNA

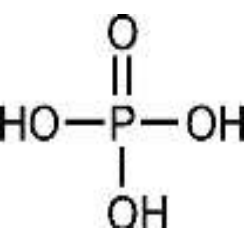
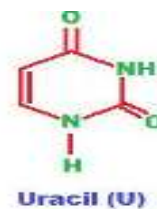
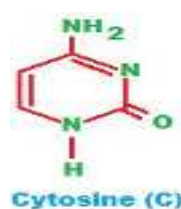
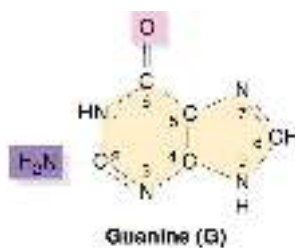
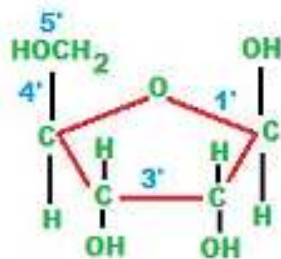
RNA (Ribonucleic acid) is a polymer of ribonucleotides.

Each ribonucleotide is composed of three components:

A ribose sugar

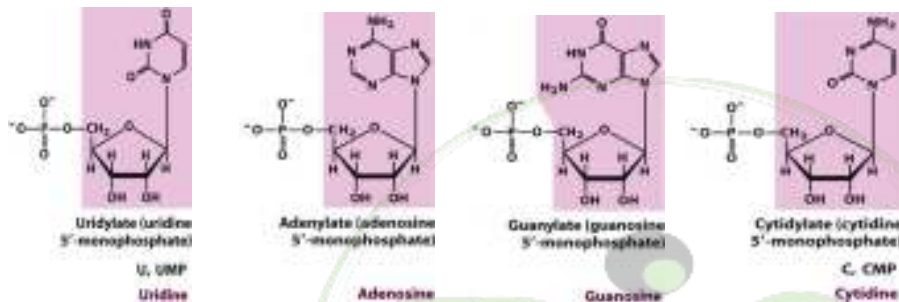
A Nitrogenous Base

A Phosphoric acid

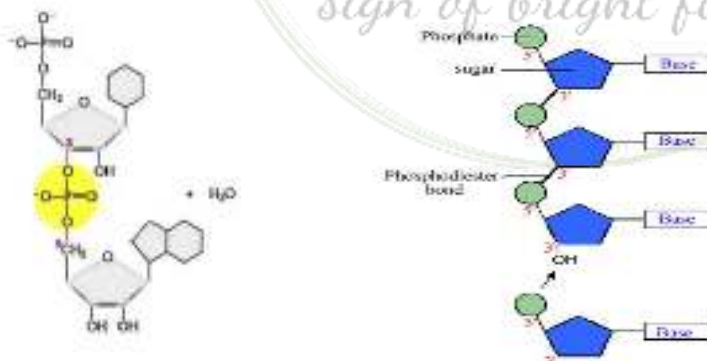


18 Types of Ribonucleotides

There are mainly four types of ribonucleotides depending upon the types of nitrogenous bases present in RNA.



How do Ribonucleotides Join? a Poly-Ribonucleotide



19 Types of RNAs

There are mainly three types of Ribonucleic acids (RNAs) present in the cells of living organisms.

Messenger RNA (mRNA)

Transfer RNA (tRNA)

Ribosomal RNA (rRNA)

Messenger RNA (mRNA)

It is the type of RNA that carries genetic information from DNA to the protein biosynthetic machinery of the ribosome.

It provides the templates that specify amino acid sequences in polypeptide chains.

The process of forming mRNA on a DNA template is known as transcription.

It may be monocistronic or polycistronic.

The length of mRNA molecules is variable and it depends on the length of gene.

Transfer RNA (tRNA)

Transfer RNAs serve as adapter molecules in the process of protein synthesis.

They are covalently linked to an amino acid at one end.

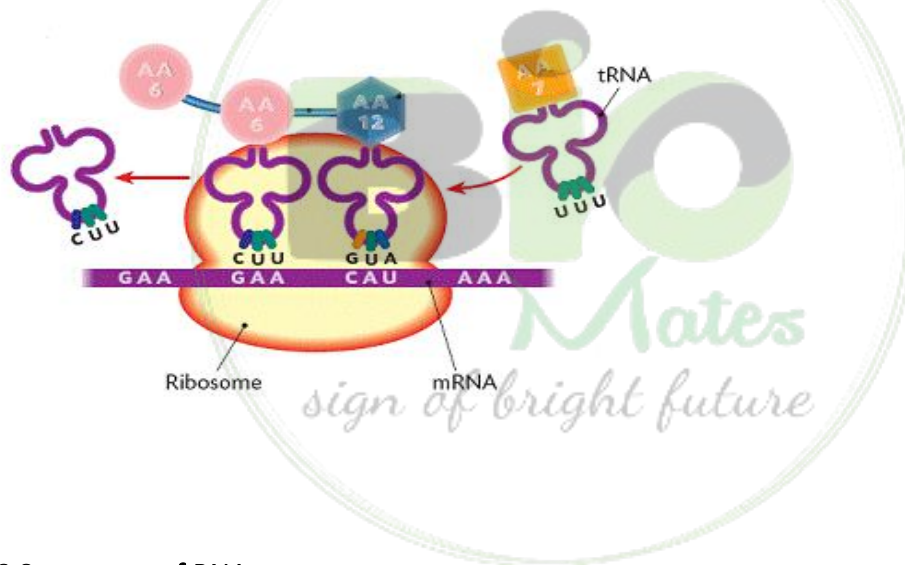
They pair with the mRNA in such a way that amino acids are joined to a growing polypeptide in the correct sequence.

Ribosomal RNA (rRNA)

Ribosomal RNAs are components of ribosomes.

rRNA is a predominant material in the ribosomes constituting about 60% of its weight.

It has a number of functions to perform in the ribosomes.



20 Structures of RNAs

mRNA is always single stranded when it is formed from DNA.

But this single strand assumes a double helical conformation soon after its formation.

This confirmation is achieved mainly due to base stacking interactions.

Messenger RNA (mRNA)



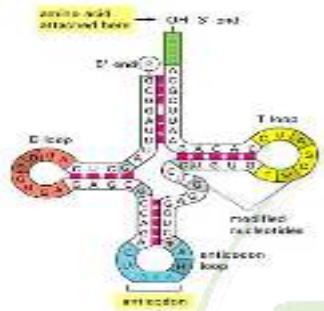
Self-complementary sequences may occur in the RNA molecules which produce more complex structures.

So RNA can base-pair with complementary regions of either RNA or DNA.

RNA has no any regular secondary structure that serves as a reference point. The three-dimensional structures of many RNAs are complex and unique.

Breaks in the helix caused by mismatched or unmatched bases in one or both strands are common and result in bulges or internal loops.

Hairpin loops form between nearby self-complementary sequences.



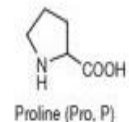
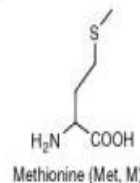
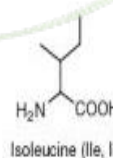
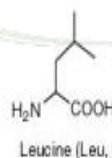
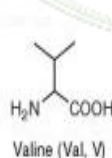
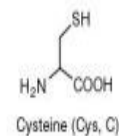
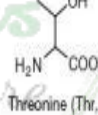
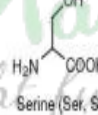
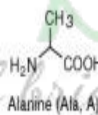
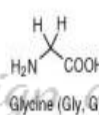
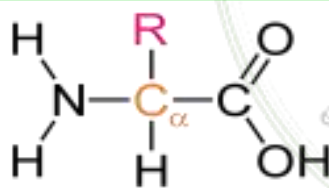
transfer RNA

21 Chemical composition of proteins

Proteins are polymers of amino acids.

They range in size from small to very large.

All the proteins are made up of Twenty different types of amino acids. So these amino acids are called standard a



In a protein molecule, each amino acid residue is joined to its neighbour

by a specific type of

covalent bond

which is called

Peptide Bond.

Amino acids can

successively join to

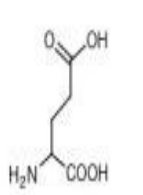
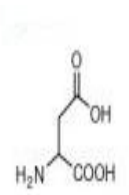
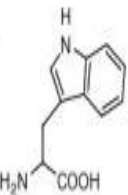
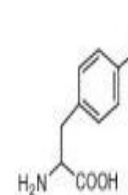
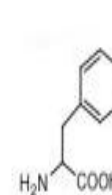
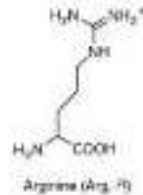
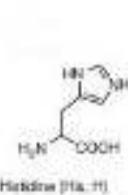
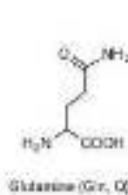
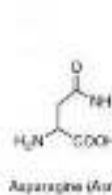
form dipeptides,

tripeptides,

tetrapeptides, oligo

peptides and

polypeptides.



22 Primary structure of proteins

Primary structure or covalent structure of protein refers to the amino acid sequence of its polypeptide chain.

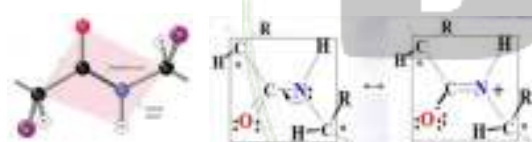
Each type of protein has a unique amino acid sequence.

Linus Pauling and Robert Corey carefully analyzed the peptide bond.

Their findings laid the foundation for our present understanding of protein structure.

They demonstrated that the peptide C - N bond is somewhat shorter than the C - N bond in a simple amine.

The six atoms of the peptide group are co-planar i.e., lie in a single plane, with the oxygen atom of the carbonyl group and the hydrogen atom of the amide nitrogen trans to each other.



Pauling and Corey concluded that the peptide C - N bonds are unable to rotate freely because of their partial double-bond character.

Rotation is permitted about the N - C and the C - C bonds.

The bond angles resulting from rotations at C are labelled ϕ (phi) for the N - C bond and ψ (psi) for the C - C bond.

In principle, ϕ and ψ can have any value between +180 & -180

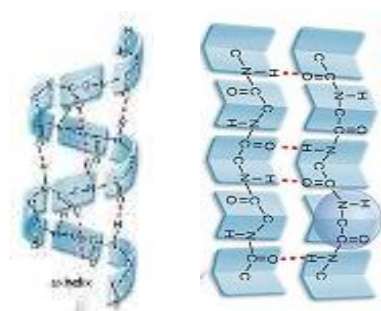
23 Secondary structure of proteins

Secondary structure of proteins refers to the local conformation of some part of a polypeptide.

A few types of secondary structures are particularly stable and occur widely in proteins.

The most prominent are:-

- α -helix
- β - conformations.



24 α - Helix

The simplest arrangement which a polypeptide chain could assume with its rigid peptide bonds is a helical structure, which Pauling and Corey called the α -helix. The helical twist of the α -helix found in all proteins is right-handed.

The repeating unit is a single turn of the helix, which extends about 5.4 Å (includes 3.6 amino acid

residues) along the long axis.

The amino acid residues in an α helix have conformations with $\psi = -45^\circ$ to -50° and $\phi = -60^\circ$.

An α helix makes optimal use of internal hydrogen bonds.

About one-fourth of all amino acid residues in polypeptides are found in α -helices while in some proteins it is the predominant structure.

25 β - Pleated Sheets

Pauling and Corey predicted a second type of secondary structure which they called β -sheets.

This is a more extended conformation of polypeptide chains.

The backbone of the polypeptide chain is extended into a zigzag structure.

The zigzag polypeptide chains are arranged side by side to form a structure resembling a series of pleats.

The R groups of adjacent amino acids protrude from the zigzag structure in opposite directions.

Hydrogen bonds are formed between adjacent segments of polypeptide chain.

The adjacent polypeptide chains in a β sheet can be either parallel or antiparallel.



26 Tertiary Structure of Proteins

The overall three-dimensional arrangement of all atoms in a protein is referred to as the protein's tertiary structure.

It includes longer-range aspects of amino acid sequence.

Amino acids that are far apart in the polypeptide chain may interact within the completely folded structure of a protein.

Interacting segments of polypeptide chains are held in their characteristic tertiary positions by different kinds of weak interactions (and sometimes by covalent bonds) between the segments.

Large polypeptide chains usually fold into two or more globular clusters known as domains, which often give these proteins a bi- or multilobal appearance.



27 Quaternary Structure of Proteins

Some proteins contain two or more separate polypeptide chains or subunits, which may be identical or different.

The spatial arrangement of these subunits is known as a protein's quaternary structure.

A multi-subunit protein is also referred to as a multimer.

A multimer with just a few subunits is called as oligomer and a single subunit or a group of subunits, is called a protomer.

Identical subunits of multimeric proteins are generally arranged in a symmetric patterns.

Oligomers can have either rotational symmetry or helical symmetry.

Quaternary Structure of Proteins

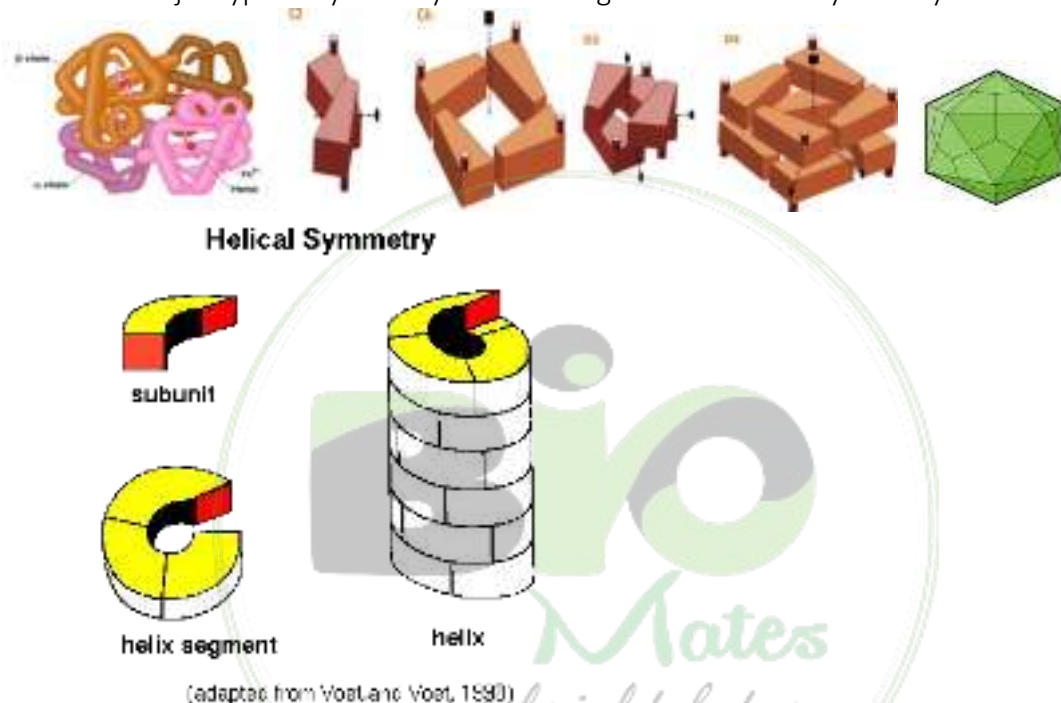
There are several forms of rotational symmetry. The simplest is cyclic symmetry, involving rotation about a single axis.

A somewhat more complicated rotational symmetry is dihedral symmetry, in which a twofold rotational axis is present.

More complex rotational symmetries include icosahedral symmetry.

An icosahedron is a regular 12-cornered polyhedron having 20 triangular faces.

The other major type of symmetry found in oligomers is helical symmetry.



28 Genetic Materials in Viruses

Viruses are exceptionally simple and extremely small microorganisms.

They have a very simple structural organization consisting of a molecule of nucleic acid and a protein coat.

Viruses can have either DNA or RNA as a genetic material but never both.

The nucleic acid of a virus can be single-stranded or double-stranded.

There are viruses with the familiar double-stranded DNA, with single-stranded DNA, with double-stranded RNA and with single-stranded RNA.

The viral nucleic acid can be linear or circular.

Some viruses may contain coiled RNA.

In some viruses, the nucleic acid is in several separate segments.

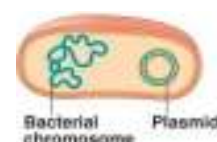
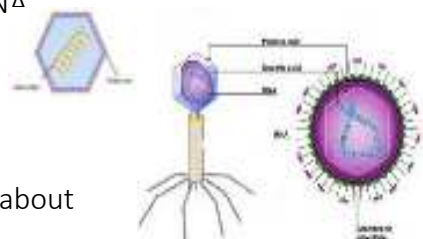
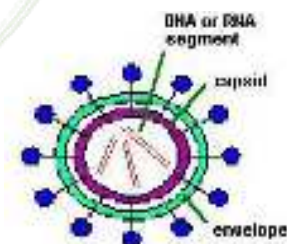
The percentage of nucleic acid in relation to protein is about 1% for the influenza virus and about 50% for some bacteriophages.

The total amount of nucleic acid varies from a few thousand nucleotides to as many as 250,000 nucleotides.

E. coli's chromosome consists of approx. 4 million nucleotide pairs.

29/30/31 Organization of Genetic Material in Bacteria

Bacteria typically have a single circular chromosome consisting of a single circular molecule of DNA with associated proteins.



The bacterial chromosome is a very long (up to 1mm).

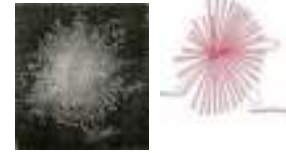
It is looped and folded and attached at one or several points to the plasma membrane.

Specific proteins interact with the bacterial DNA to form a highly condensed nucleoprotein complex called the nucleoid.



Bacterial chromatin can be released from the cell by gentle lysis of the cell.

Electron micrograph of the chromatin reveals that it consists of multiple loops which emerge from a central region of the chromatin.



Some of the loops are super-coiled while some are relaxed.

Relaxed loops are formed as a result of a nick introduced into super-coiled loops by a cellular DNase.

If a super-coiled DNA molecule receives a nick, the strain of under-winding is immediately removed, and all the super-coiling is lost.

Studies confirm that continued nuclease treatment increases number of relaxed loops.

The bacterial DNA is arranged in super-coiled loops that are fastened to a central protein matrix, so that each loop is topologically independent from all the others. So a nick that causes one super-coiled loop to relax would have no effect on other super-coiled loops.

The super-coiled loops are dynamic structures which change during cell growth & division.

An E. coli chromosome is estimated to have about 400 super-coiled loops.

Each loop has an average length of about 10-20 kbp.

The DNA compaction in a bacterial cell is contributed by super-coiling of loops, macromolecular crowding and DNA-binding proteins.

32 Genetics and Genomics

Genetics and Genomics - Difference

Genetics

Genetics is the study of heredity, or how the characteristics of living organisms are transmitted from one generation to the next generation through DNA.

Genetics involves the study of specific and limited numbers of genes that have a known function.

Genetics deals that how genes guide the body's development, cause disease or affect response to drugs.

Genomics

Genomics, in contrast, is the study of the entirety of an organism's genes – called the genome.

Using high-performance computing and math techniques known as bioinformatics, genomics analyzes enormous amounts of DNA sequence data to find variations.

Genomics particularly deals with genetic variants that affect health, disease or drug response.

In humans that means searching through about 3 billion units of DNA across 23,000 genes.

Genomics is a much newer field than genetics and became possible only in the last couple of decades due to technical advances in DNA sequencing and computational biology.

Genetics: How the characteristics of living organisms are transmitted from one generation to the next generation.

Genomics: study of the entirety of an organism's genes – called the genome.

33 Genomics, Proteomics and Metabolomics

Genomics, Proteomics, Metabolomics

Genome and Genomics

The complete set of DNA found in each cell is known as the genome and study is called as genomics.

Proteome and Proteomics

The complete set of proteins found in each cell is known as the proteome.

Proteins concentration (and activity) may be different than gene expression due to post-translational modification

Metabolomics

The complete set of metabolites found in each cell is known as the metabolome.

Use of high-throughput mass spectrometry to analyze the metabolic components of cell.

Useful for determining the effects of the environment or gene transformation on the metabolism of the plants/animals.

Conclusion

Genomics, proteomics and metabolomics will give an integrated, wholistic view of the cell.

Can be used to monitor or modify organisms in a comprehensive way.

Bioinformatics - the key to understand the plethora of information and modeling the cell.

34 Why Sequence Genomes

To identify gene numbers, their locations on genomes, and to study their functions.

Genes regulation

DNA sequence

Genome organization

Chromosomal structure and organization

Noncoding DNA types, amount, distribution and functions.

Coordination of gene expression, protein synthesis, and post-translational events.

Interaction of proteins in complex molecular machines

Predicted vs experimentally determined gene function

Evolutionary conservation

- Proteins structure and function.
- Proteomes (total protein content and function) in organisms.
- Correlation of SNPs with health and disease

Disease-susceptibility prediction based on gene sequence variation
Genes involved in complex traits and multigene diseases

Novel Diagnostics

Complex systems biology, developmental genetics.
To provide platform for microchips and DNA microarrays.
Gene expression - RNA
Complex systems biology, developmental genetics and genomics
Drug target discovery
Rational drug design
Molecular docking
Gene therapy
Stem cell therapy

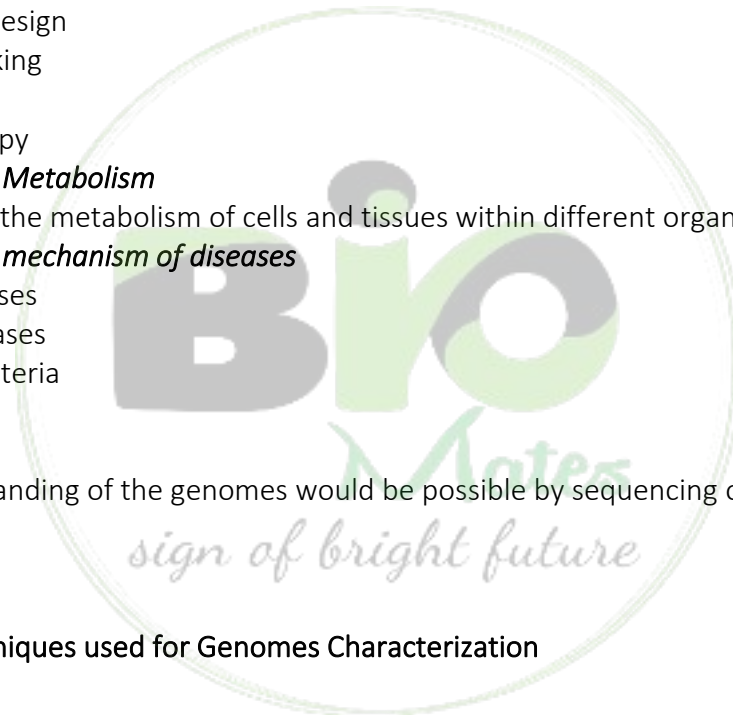
Understanding Metabolism

To understand the metabolism of cells and tissues within different organisms.

Understanding mechanism of diseases

Inherited diseases
Infectious diseases
Pathogenic bacteria
Viruses
Conclusions

Better understanding of the genomes would be possible by sequencing of the genomes.



35 Major Techniques used for Genomes Characterization

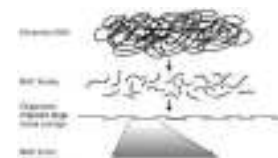
Genomes Characterization -Techniques

Major Techniques used for Genomes Characterization

Cloning
Hybridization
PCR amplification
Sequencing
Computational tool
Genomes digested with restriction enzymes and inserted in vectors to produce genomic libraries.
BACs
YACs

Genomes Characterization Techniques - Hybridization

To arrange large contigs of genomes to produce genetic maps and physical maps of genomes. To arrange large contigs of genomes to produce genetic and physical maps



Genomes Characterization Techniques – PCR

Technique to amplify the DNA. Different variants of the technique used

Genomes Characterization Techniques – DNA Sequencing

One of the important technique used to characterize the genomes

To study structure and function of genomes.

DNAs are amplified and sequenced

Genomes Characterization Techniques – Computational Tools

Used to align the sequenced DNA to produce physical maps of the genomes.

Genomes Characterization Techniques – Conclusion

Different techniques used for genomes characterizations.

36 Genomes Analysis - Steps

Steps of Genomes Analysis

- Genome sequence assembled
- Identify repetitive sequences – mask out
- Gene prediction – train a model for each genome
- Look for EST and cDNA sequences
- Genome annotation
- Microarray analysis
- Metabolic pathways and regulation
- Protein 2D gel electrophoresis
- Functional genomics
- Gene location/gene map
- Self-comparison of proteome
- Comparative genomics
- Identify clusters of functionally related genes
- Evolutionary modeling

37 Benefits of Genomes Research

Genomes Research – Molecular Medicine

Improve diagnosis of disease

Detect genetic predispositions to disease (cancer, diabetes etc)

Create drugs based on molecular information

Use gene therapy and control systems as drugs

Genomes Research – Risk Assessment

Evaluate the health risks faced by individuals who may be exposed to radiations and to cancer causing chemicals and toxins.

Genomes Research – Bioarchaeology, Anthropology, Evolution and Human Migration

Study evolution through genetic variants in lineages.

Study of migration of different populations

Study mutations on the Y chromosome to trace lineage and migration of males

Evolution of mutations with ages of populations

Genomes Research – DNA Forensics

Identify potential suspects whose DNA may match evidence left at crime scenes.
Exonerate persons wrongly accused of crimes.
Identify catastrophe victims.
Establish paternity and other family relationships.
Identify endangered and protected species as an aid to wildlife officials.

Genomes Research

Detect bacteria and other organisms that may pollute air, water, soil and food.
Match organ donors with recipients in transplant programs
Determine pedigree for seed or livestock

Genomes Research – Disease-resistant crops and disease-resistant animals

Grow disease/insect resistant and drought-resistant crops.
Breed healthier, more productive, disease-resistant farm animals.
Develop biopesticides.
Incorporate edible vaccines incorporated into food products.

Genomes Research – Microbial Genomics

Rapidly detect and treat pathogens (disease-causing microbes).
Develop new energy sources (biofuels)
monitor environment to detect pollutants.
Protect populations from biological and chemical warfare
Clean up toxic waste safely and efficiently

38 Size of Genomes

Genomes of most bacteria and archaea range from 1 to 6 million base pairs (Mb).
Genomes of eukaryotes are usually larger

Size of Genomes

Most plants and animals have genomes greater than 100 Mb.
Humans have genome size of 3,000 Mb

Within each domain there is no systematic relationship between genome size and phenotype

Organism	Haploid Genome Size (Mb)	Number of Genes	Genes per Mb
Bacteria			
<i>Haemophilus influenzae</i>	1.8	1,700	940
<i>Escherichia coli</i>	4.6	4,400	950
Archaea			
<i>Archaeoglobus fulgidus</i>	2.2	2,500	1,130
<i>Methanosarcina barkeri</i>	4.8	3,600	750

Organism	Haploid Genome Size (Mb)	Number of Genes	Genes per Mb
Eukaryotes			
<i>Saccharomyces cerevisiae</i> (yeast, a fungus)	12	6,300	525
<i>Caenorhabditis elegans</i> (nematode)	100	20,100	200
<i>Arabidopsis thaliana</i> (mustard family plant)	120	27,000	225
<i>Drosophila melanogaster</i> (fruit fly)	165	13,700	83
<i>Oryza sativa</i> (rice)	430	42,000	98
<i>Zea mays</i> (corn)	2,300	32,000	14
<i>Mus musculus</i> (house mouse)	2,600	22,000	11
<i>Ailuropoda melanoleuca</i> (giant panda)	2,400	21,000	9
<i>Homo sapiens</i> (human)	3,000	<21,000	7

Conclusion

Although most eukaryotes have large size of genomes.

Within each domain there is no systematic relationship between genome size and phenotype

39 viral Genomes

Genomes of Viruses

Viral genomes can be

- ssRNA
- dsRNA
- ssDNA
- dsDNA
- Linear
- Circular

In Virus

- A viral genome is the genetic material of the virus.
- Also termed the viral chromosome.
- Viral genomes vary in size -few thousand to more than a hundred thousand nucleotides.

Virus with RNA

- Viruses with RNA Genomes
- Almost all plants viruses and some bacterial and animal viruses
- Genomes are rather small (a few thousands nucleotides)

Viruses with DNA Genomes

- Often a circular genome
- lambda = 48,502 bp

Replicative form of Viral Genomes

- All ssRNA viruses produce dsRNA molecules
- Many linear DNA molecules become circular

Viruses and Kingdoms

- Many plants viruses contain ssRNA genomes.
- Many fungal viruses contain dsRNA genomes.
- Many bacterial viruses contain dsDNA genomes.

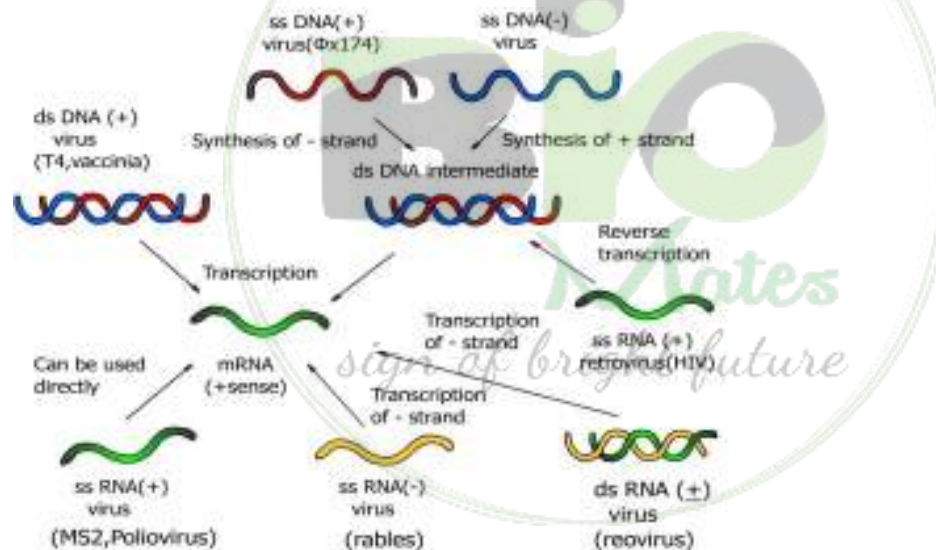
Genomes in Virions: The genomes of viruses can be composed of either DNA or RNA, and some use both as their genomic material at different stages in their life cycle. However, only one type of nucleic acid is found in the virion of any particular type of virus.

Virus and number of Gene

Virus	Host	Type of Nucleic Acid	Number of Genes
Parvovirus	Mammals	ssDNA	5
Phage fd	<i>E. coli</i>	ssDNA	10
Lambda	<i>E. coli</i>	dsDNA	36
T4	<i>E. coli</i>	dsDNA	>190
Q β	<i>E. coli</i>	ssRNA	4
TMV	Many plants	ssRNA	6
Influenza virus	Mammals	ssRNA	12

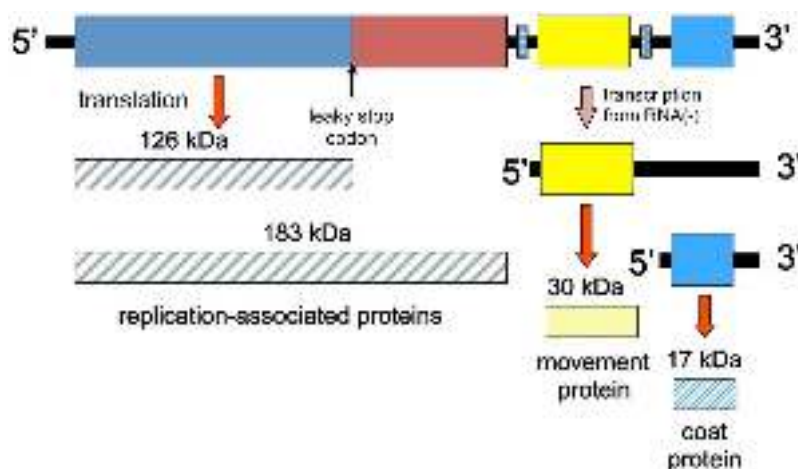
Virus	Genome structure	Genome size (kb)	Number of genes
Adenovirus	Double-stranded linear DNA	36.0	30
Hepatitis B	Partly double-stranded circular DNA	3.2	4
Influenza virus	Single-stranded segmented linear RNA	22.0	12
Parvovirus	Single-stranded linear DNA	1.6	5
Poliovirus	Single-stranded linear RNA	7.6	8
Reovirus	Double-stranded segmented linear RNA	22.5	22
Retroviruses	Single-stranded linear RNA	6.0–9.0	3
SV40	Double-stranded circular DNA	5.0	5
Tobacco mosaic virus	Single-stranded linear RNA	6.4	6
Vaccinia virus	Double-stranded circular DNA	240	240

Genomes in Virions



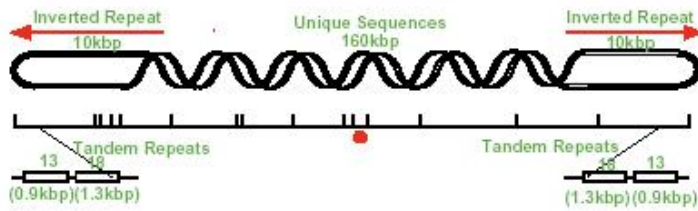
Genome of Tobacco Mosaic Virus

Single, 6400 nucleotides RNA, 3 Essential Genes

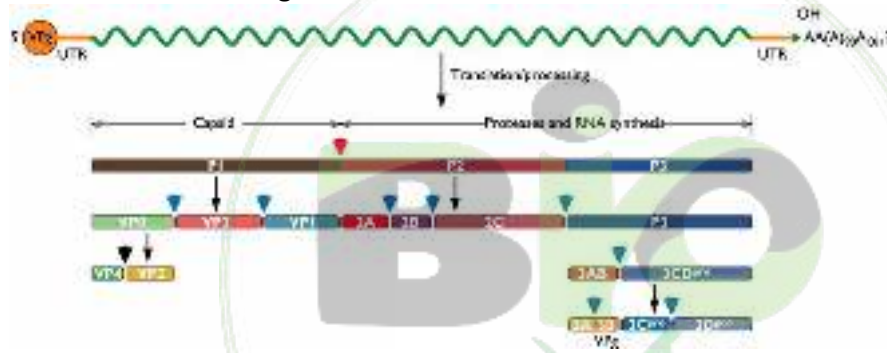


Genome of Poxvirus –A typical large dsDNA Virus

180 kb DNA, >100 Essential Genes



Genome of Polio Virus: Single-stranded positive-sense RNA genome that is about 7500 nucleotides long



Genome of Pox Virus

Linear dsDNA 130-375 kbp; covalently closed termini.

Large hairpin structure at each terminus - up to 10 kb total at each end is repeat sequence.

Encode 150-300 proteins.

Coding regions are closely spaced, no introns.

Coding regions are on both strands of genome, and are not tightly clustered with respect to time of expression or function.

Viral genomes can be

- ssRNA
- dsRNA
- ssDNA
- dsDNA
- Linear
- Circular

40 Bacterial Genomes

Small organisms carry high coding density (85-90%)

1 gene per 1000 bases in prokaryotes

Large variation in genome size between bacteria

Genomes of Bacteria – Large Variation

Tremblaya princeps 140kb, 121 coding

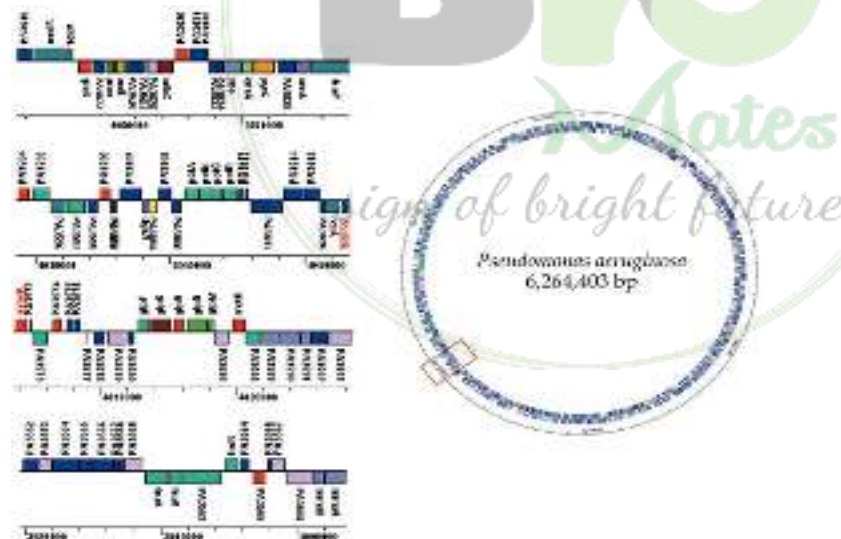
Microorganism	# Genes in the Genome	# Regulatory Proteins	% of Total
<i>Pseudomonas aeruginosa</i>	5570	468	8.4
<i>Escherichia coli</i>	4289	290	5.8
<i>Bacillus subtilis</i>	4100	217	5.3
<i>Mycobacterium tuberculosis</i>	3918	117	3.0
<i>Helicobacter pylori</i>	1566	18	1.1

sequences
Sorangium cellulosum
14000kb
11599 coding sequences

[illegible]

Distribution of genes among selected bacterial genomes and their sizes

Genes in a portion of a bacterial genome



Bacterial Genomes - Conclusion

- Small organisms carry high coding density.
- Large variation in genome size between bacteria.

41 Yeast Genome

The nuclear genome consists of 16 chromosomes.
In addition, there is a mitochondrial genome and a plasmid, 2 micron circle.

The haploid yeast genome consists of ~ 12.1 Mb
Yeast genome was completely sequenced by 1996

Yeast Genome - Characteristics

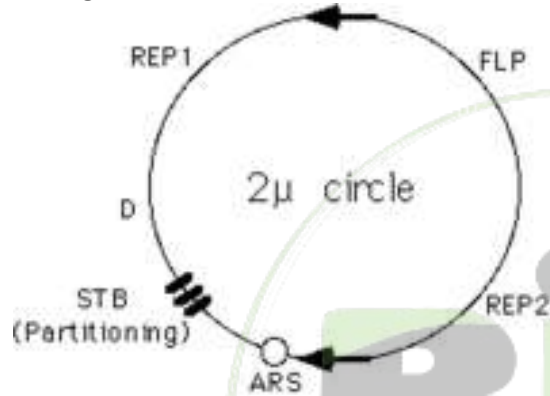
- Small and compact
- Small intergenic sequences

The 2 μ circle is a 6.3 kb

50 to 100 copies per haploid genome of the yeast cells

ARS, the FLP gene, the three genes which encode proteins required for regulation of FLP expression (REP2, REP1, and D)

Set of small direct repeats (called "STB") required for partitioning into daughter cells during mitosis and meiosis.



Yeast nuclear genome has 16 chromosomes.

A mitochondrial genome.

A plasmid.

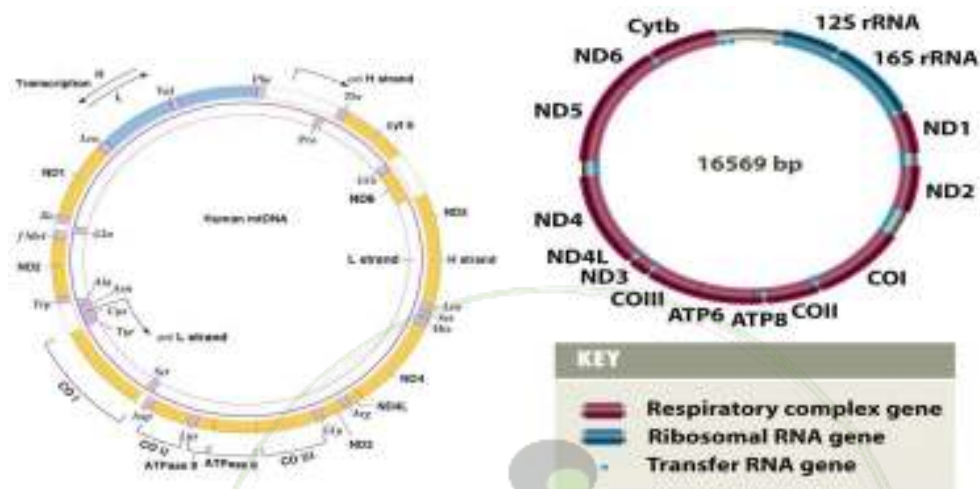
42 Mitochondrial Genome

- Multiple identical circular chromosomes
- ~15-16 Kb in animals
- ~ 200 kb to 2,500 kb in plants
- Over 95% of mitochondrial proteins are encoded in the nuclear genome.
- Often A+T rich genomes.

Human Mitochondrial Genome

- Circular, double stranded, 16.6 kb
- The two strands are notably different in base composition, leading to one strand being heavy (H strand) and the other light (L strand).
- Both strands encode genes, although more are on the H strand.
- A short region (1121 bp), the D loop is a DNA triple helix: two overlapping copies of the H strand.
- The D loop is also the site where most of replication and transcription is controlled.
- Genes are tightly packed, with almost no non-coding DNA outside of D loop.
- Human mitochondrial genes contain no introns, although introns are found in the mitochondria of other groups (plants)
- 37 Genes
- 22 tRNAs
- 2 rRNAs
- 13 polypeptides
- tRNA: only 60 of the 64 codons code for amino acids.

Mitochondrial Genome Human



Mitochondrial Genome – Yeast



Mitochondrial Genome – Human and Yeast

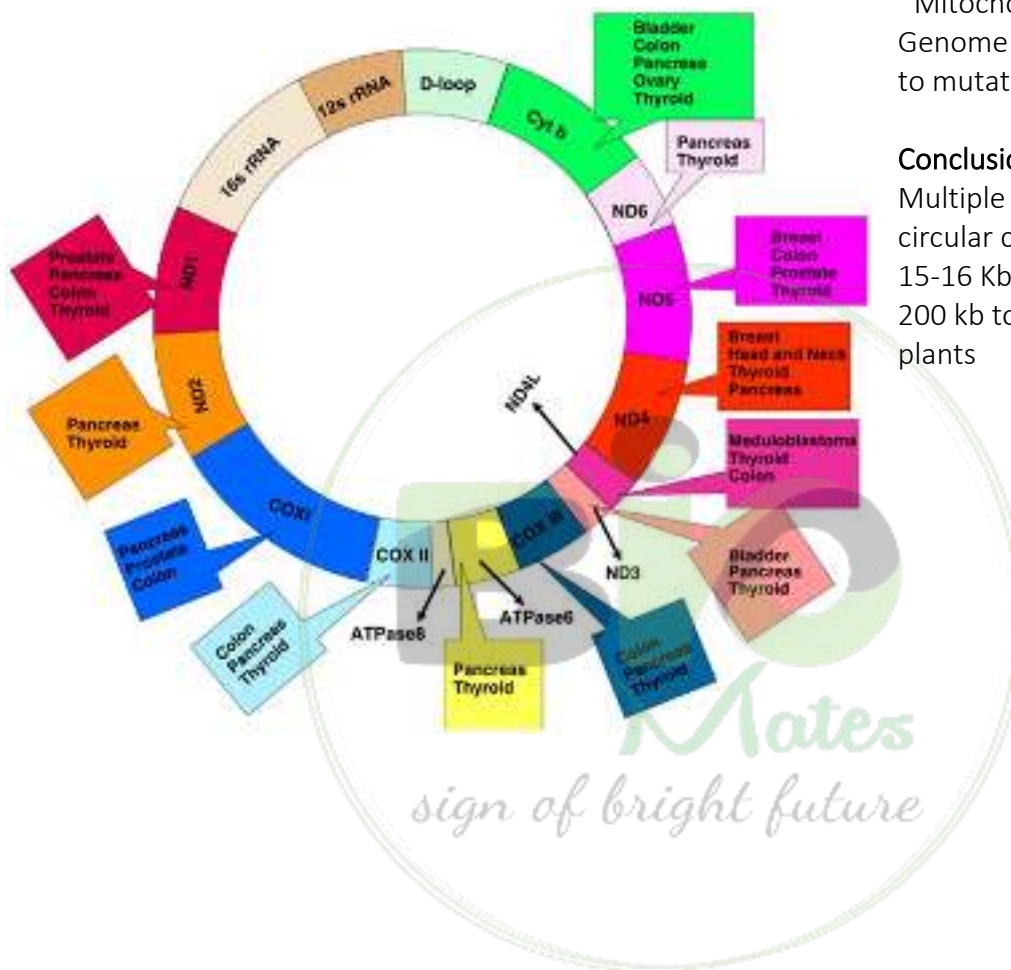


Gene	Protein-coding genes	Non-coding introns	Flank genes	Non-coding proteins	Non-coding RNAs	Non-coding overlap
Total number of genes	5	12	17	10	53	93
Types of genes						
Protein-coding genes	5	2	18	8	27	63
Mitochondrial complex	5	2	15	5	13	24
Non-coding proteins	0	0	0	3	3	27
Non-coding proteins	0	0	0	0	0	0
RNA polymerase	0	0	0	0	0	4
Non-coding factor	0	0	0	0	0	1
Functional RNA genes	2	5	24	17	28	50
Non-coding RNA genes	2	2	2	2	2	3
Non-coding RNA genes	0	3	22	24	22	28
Other RNA genes	0	0	0	1	0	1
Number of RNAs	0	1	0	8	23	1
Genome size (kb)	5	18	17	75	307	69

Universal code

		Second base								
		U		C		A		G		
First base	U	UUU	Phe	UUC		UAU	Tyr	UGU	Cys	U
		UUA		UUG	Ser	UAC		UGC		C
		UUA	Leu	UGA		UAA	Stop	UGG	Stop	G
	C	CUU		CCU		CAU	His	CGU	Arg	U
		CUC	Leu	CCC	Pro	CAC		CGC		C
		CUA		CCA		CAA	Gln	CGA	Arg	A
	A	AUU		ACU		AAU		AGU	Arg	U
		AUA	Ile	ACC		AAC	Ser	AGC		C
		AUG	Met	ACA	Thr	AAA		AGA	Arg	A
	G	GUU		GGU		GAU	Asp	GGU	Arg	U
		GUC	Val	GGC		GAC		GGC		C
		GUA		GGA	Ala	GAA	Glu	GGA	Gly	A

Universal Code and Mitochondrial Code



Mitochondrial
Genome – Tumors due
to mutations

Conclusion

Multiple identical
circular chromosomes
15-16 Kb in animals
200 kb to 2,500 kb in
plants

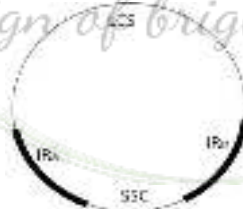
43 Chloroplast Genome

- Multiple circular molecules
- Size ranges from 70 kb to 2000 kb
- Land plants typically 120 – 170 kb
- ~ 70 kb – Epifagus
- ~2,000 kb – Acetabularia
- Similar to mtDNA
- Many chloroplast proteins are encoded in the nucleus (separate signal sequence)
- Double stranded DNA molecule
- Chloroplasts genomes are relatively larger
- Multiple copies of genome
- Large enough to code 50-100 proteins as well as rRNAs & tRNAs
- cpDNA regions includes Large Single Copy & Small Single-Copy (SSC) regions, and Inverted Repeats (IRA & IRB).

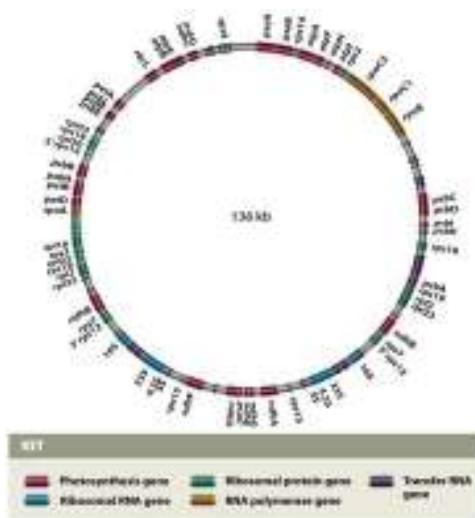
Taxa	Genome size (n kb)	Inverse duplication (n kb)
Angiosperms		
<i>Nicotiana glauca</i>	158	25
<i>Solanum elaeagnifolium</i>	150	26
<i>Antirrhinum majus</i>	217	28
<i>Arabidopsis thaliana</i>	120	Not present
<i>Populus nigra</i>	70	27
<i>Oryza sativa</i>	134	21
Gymnosperms		
<i>Pinus</i>	120	Not present
<i>Quercus ilex</i>	158	27
Peridophyta		
<i>Chara corallina</i>	144	10
Bryophyta		
<i>Marchantia polymorpha</i>	121	10
Chlorophyta		
<i>Chlamydomonas reinhardtii</i>	85	Not present
<i>Chlamydomonas reinhardtii</i>	195	24
<i>Chlamydomonas reinhardtii</i>	242	41
Rhodophyta		
<i>Chlorophyta reinhardtii</i>	127	12
Chromophyta		
<i>Chlamydomonas reinhardtii</i>	123	5

Species	Type of organism	Genome size (kb)
Chloroplast genomes		
<i>Pisum sativum</i>	Flowering plant (pea)	120
<i>Marchantia polymorpha</i>	Liverwort	121
<i>Oryza sativa</i>	Flowering plant (rice)	136
<i>Nicotiana glauca</i>	Flowering plant (tobacco)	156
<i>Chlamydomonas reinhardtii</i>	Green alga	195

- Large Single Copy
- Small Single Copy
- Inverted Repeats



Region
Region



Chloroplast Genome of Rice Functions of Chloroplast Genes

- Most cp genes fall into two functional groups:
- Genes involved in replication, transcription, translation
- Genes involved in photosynthesis

Genes Nomenclature

- Based on bacterial naming system, which uses lower case letters, and a descriptive prefix, based on the probable function. If the gene product is part of a multi-subunit complex, a letter of the

alphabet is used to denote different subunits.

- psa for genes of photosystem I (psaA, psaB, etc.)
- psb for genes of photosystem II (psbA, psbB, etc.)
- Non-mendelian inheritance
- Self replication

- Somatic segregation in plants
- Inherited independently of nuclear genes
- Conservative rate of nucleotide substitution enables to resolve plant phylogenetic relationships at deep levels of evolution.
- Multiple circular molecules
- Size ranges from 70 kb to 2000 kb

44 Eukaryotic Genomes

Located on several chromosomes

Relatively low gene density

Carry organelles genome in addition to nuclear genome

Contains repetitive sequences

SINEs (short interspersed elements)

LINEs (long interspersed elements)

LINES and SINES

LINES

1-5 kb

10-10,000 copies

SINES

200-300 bp

100,000 copies

Highly repetitive

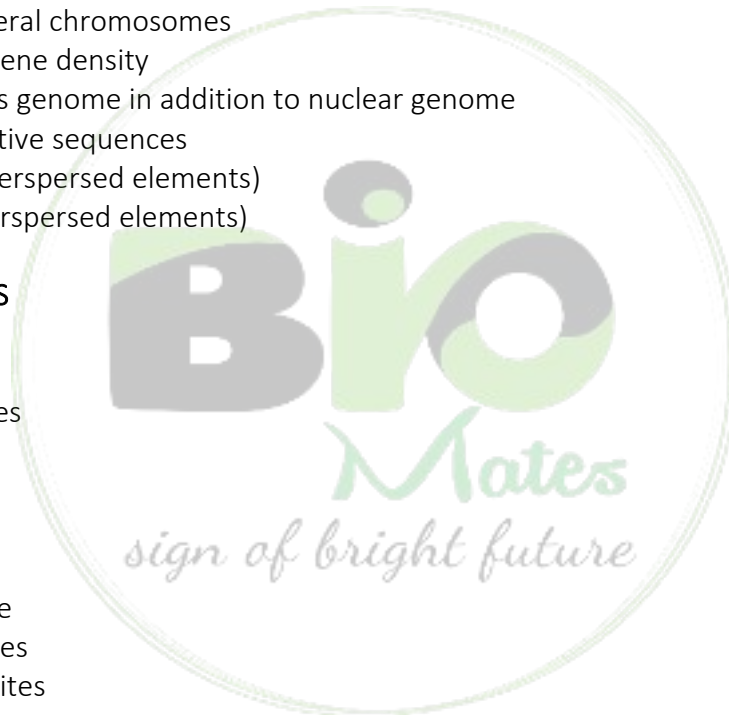
- Minisatellites
- Microsatellites
- Telomeres

Minisatellites

- Repeats of 14-500 bp
- 1-5 kb long
- Scattered throughout the genome
- Repeats up to 13 bp

Telomeres

- Short repeats (6 bp)
- 250-1,000 at ends of chromosomes



Chromosomes: linear, centromeres, telomeres, origins of replication, replicons

Protein-coding genes and spliceosomal Introns

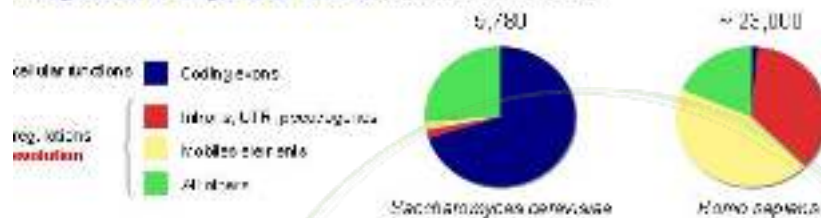
Genes for non coding RNAs: rRNAs, tRNAs, snoRNAs, snRNAs, microRNAs

Mobile genetic elements: and their remnants

Pseudogenes: and processed pseudogenes

Satellite DNAs: micro-, minisatellites, repeated sequences

Fragments of organellar DNAs: NUMTs and NUPTs



Prokaryotic

- Usually circular
- Smaller
- Found in the nucleoid region
- Less elaborately structured and folded

Eukaryotic

- Complex with a large amount of protein to form chromatin
- Highly extended and tangled during interphase
- Found in the nucleus

Conclusion

- Located on several chromosomes
- Relatively low gene density
- Chromosomes vary in length
- Carry organelles genome

Genomes Comparisons

Genomes vary in size

Genomes of most bacteria and archaea range from 1 to 6 million base pairs (Mb)

Most plants and animals have genomes greater than 100 Mb; humans have 3,000 Mb

Genomes Comparisons

Genomes vary in genes numbers

Free-living bacteria and archaea have 1,500 to 7,500 genes

Fungi have about 5,000 genes and multicellular eukaryotes upto 40,000 genes

Number of genes is not correlated to genome size

Nematode

C. elegans has 100 Mb and 20,000 genes, while *Drosophila* has 165 Mb and 13,700 genes

Genomes vary in genes numbers

Vertebrate genomes can produce more than one polypeptide per gene because of alternative splicing of RNA transcripts

Humans and Mammals have low gene density

Humans and other mammals have the lowest gene density, or number of genes, in a given length of DNA. Multicellular eukaryotes have many introns within genes and noncoding DNA between genes

Multicellular eukaryotes have much noncoding DNA and multigene families

Most of eukaryotic genomes neither encodes proteins nor functional RNAs

Evidence indicates that noncoding DNA plays important roles in the cell

Genomes Comparisons

Human Genome: Distribution of coding and non-coding DNA

Significant similarity between genomes of "distant" species (Man – Yeast 23%)

Similarity increases for taxonomically close species.

Closely related species help us understand recent evolutionary events

Distantly related species help us understand ancient evolutionary events

Comparing Genomes: Bacteria, archaea, and eukaryotes diverged from each other between 2 and 4 billion years ago

Human and chimpanzee genomes differ by 1.2%, at single base-pairs, and by 2.7% because of insertions and deletions

45 Comparing distantly/closely related species

Comparing distantly related species

- Highly conserved genes have changed very little over time
- These help to clarify relationships among species that diverged from each other long ago
- Bacteria, archaea, and eukaryotes diverged from each other 2 and 4 billion years ago
- Highly conserved genes can be studied in one model organism.
- Genetic differences between closely related species can be correlated with phenotypic differences
- Genetic comparison of several mammals with non-mammals helps to identify what make mammals
- Human and chimpanzee genomes differ by 1.2%, at single base-pairs, and by 2.7% because of insertions and deletions
- Several genes are evolving faster in humans than chimpanzees
- Genes involved in defense against malaria and tuberculosis and in regulation of brain size, genes code for transcription factors

- Humans and chimpanzees differ in the expression of the FOXP2 gene, whose product turns on genes involved in vocalization
- Differences in the FOXP2 gene may explain why humans but not chimpanzees communicate by speech

Conclusion

Highly conserved genes have changed very little over the time

These help to clarify relationships among species that diverged from each other long ago

47 Anatomy and Organization of Genomes

Genome Anatomy

Anatomy of different genomes differ from each other.

Eukaryotes and prokaryotes genomes differ very significantly.

Size of genomes - 1000 fold difference between eukaryotes and prokaryotes.

~ 30 fold between genomes of different eukaryotes.

In humans ~ 23,000

Bacterial genomes ~ 1,500 – 2,000 genes.

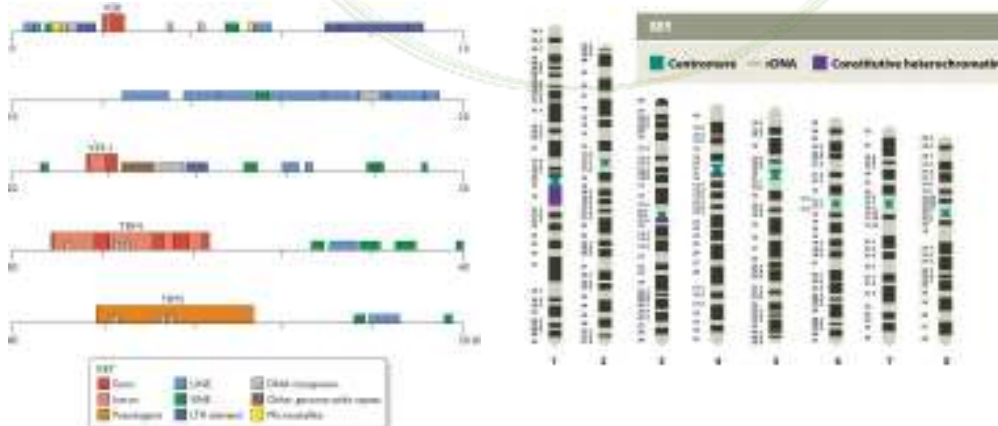
Genome Anatomy - Eukaryotes

Eukaryotic genomes are full of simple repeats, numerous types of transposable elements and other sequences.

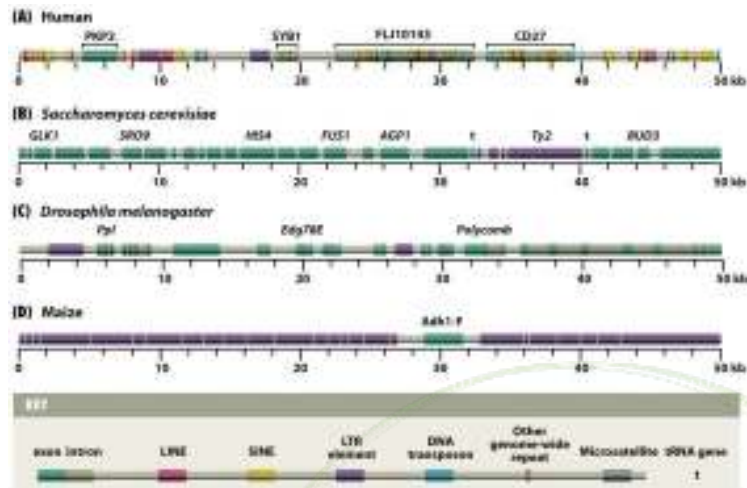
Genome Anatomy - Prokaryotes

Prokaryotes have a few repeats and transposable elements and their genomes consist of mainly the genes.

Genome Anatomy: Segment of human chromosome 7



Genome Organization: Human, Yeast, Fruit Fly , Maize



Compactness of Genomes

Conclusion

- Anatomy of different genomes differ from each other.
- Eukaryotes and prokaryotes genomes differ very significantly.

48 Gene Anatomy

What is Gene

A piece of DNA (or RNA) that contains the primary sequence to produce a functional biological gene product (RNA or protein).

Entire nucleic acid sequence necessary for the synthesis of a functional polypeptide (protein chain) or functional RNA

Genetic information is stored in DNA

Segments of DNA that encode proteins or other functional products are called genes.

Gene sequences are transcribed into messenger RNA (mRNA).

mRNA is translated into Proteins

mRNA intermediates are translated into proteins that perform most of the life functions.

Three components

Open Reading Frame: From start codon (ATG) to stop (TGA, TAA, TAG)

Upstream region with binding site. (e.g. TATA box, GC box, CAAT box).

Poly-a tail

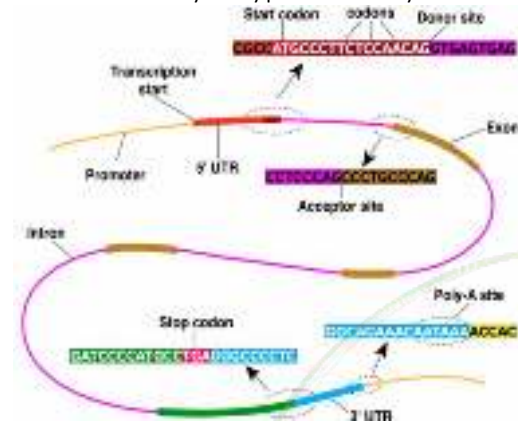
Gene Anatomy

Gene Anatomy – Typical Prokaryotic Gene



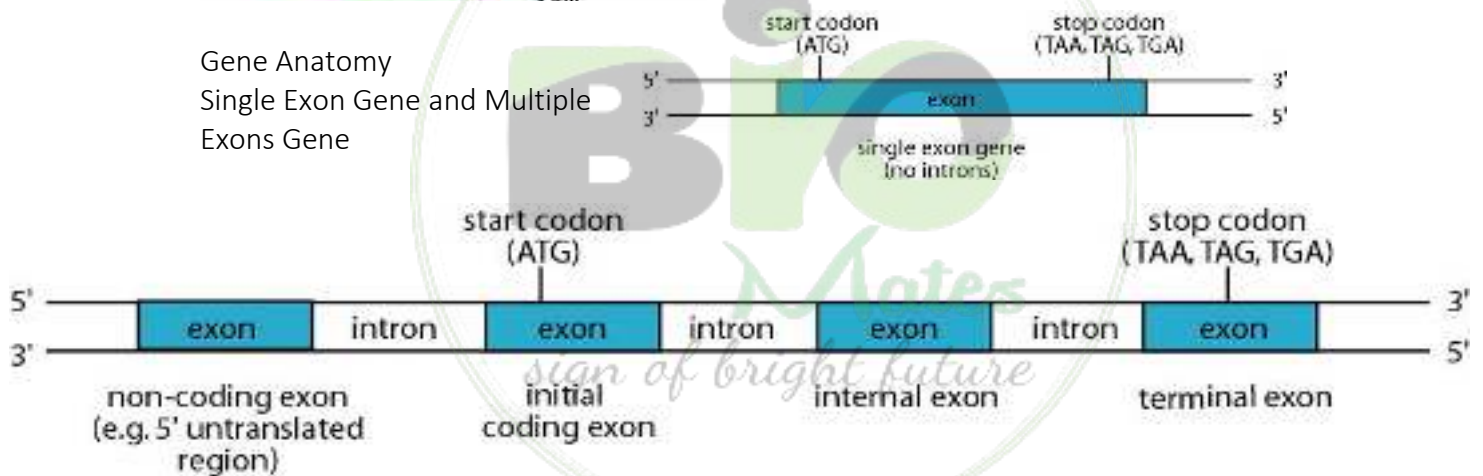
Gene Anatomy

Gene Anatomy – Typical Eukaryotic Gene



Gene Anatomy

Single Exon Gene and Multiple Exons Gene



A piece of DNA (or RNA) that contains the primary sequence to produce a functional biological gene product (RNA or protein).

49 Prokaryotic Gene and Eukaryotic Gene

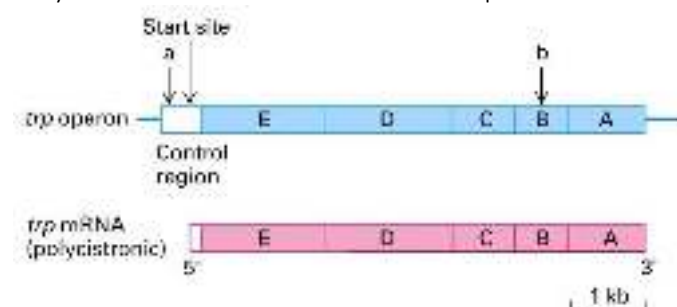
Prokaryotic Gene/Eukaryotic Gene

Bacterial Gene

Most do not have introns

Many are organized in operons: contiguous genes, transcribed as a single polycistronic mRNA, that encode proteins with related functions

Polycistronic mRNA encodes several proteins



Bacterial Gene: Polycistronic mRNA encodes several proteins

Eukaryotic Gene: Exons and Introns

Introns: intervening sequences within a gene that are not translated into a protein sequence.

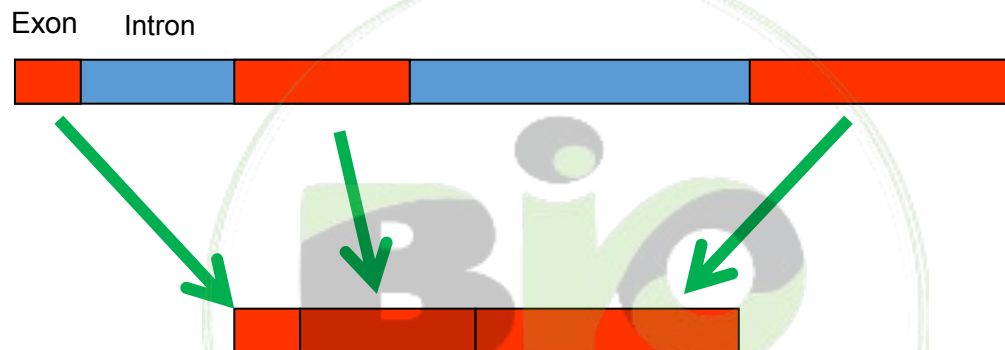
Exons: sequences within a gene that encode protein sequence.

Eukaryotic Gene

Splicing: Removal of introns from the mRNA molecule

Prokaryotic Gene/Eukaryotic Gene

Splicing: Removal of introns from the mRNA molecule



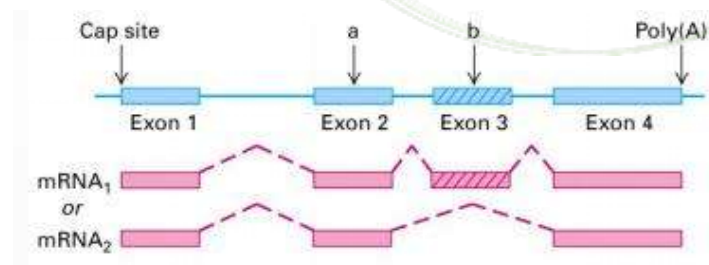
Prokaryotic Gene/Eukaryotic Gene

Organize expression of genes' (function calls)

Promoter region (binding site), usually near coding region

Binding can block (inhibit) expression

Alternative Splicing in Eukaryotic Genes



Most have introns

Produce monocistronic mRNA

Large in size

Eukaryotic Gene: Alternative Splicing

Eukaryotic Gene

Computational challenges

Identify binding sites

Correlate sequence to expression

Conclusion

Most of prokaryotic genes are without introns and in the are polycistronic.

Eukaryotic genes have introns and alternative splicing.

50 Types of Eukaryotic DNA

Protein coding genes

- Tandemly repeated genes
- Repeated DNA
- Unclassified spacer DNA
- Protein coding genes can also be in the form of:
- Duplicated and diverged genes
- Functional gene families and non-functional pseudo-genes
- Tandemly repeated genes encoding rRNA, 5sRNA, tRNA and histones.
- Repetitive DNA
- Simple sequence DNA
- Moderately repeated DNA (mobile DNA elements)
- Transposons
- Retrotransposons

Types of Eukaryotic DNA

- Repetitive DNA
- Long interspersed elements
- Short interspersed elements
- Unclassified spacer DNA

Protein-coding genes

Solitary genes

Duplicated and diverged genes (functional gene families and nonfunctional pseudogenes)

Tandemly repeated genes encoding rRNA, 5S rRNA, tRNA, and histones

Repetitious DNA

Simple sequence DNA

Moderately repeated DNA (mobile DNA element)

Transposons

Viral retrotransposons

Long interspersed elements (LINES; nonviral retrotransposons)

Short interspersed elements (SINES; nonviral retrotransposons)

Unclassified spacer DNA

Class	Length	Copies/Number in Human Genome	Proportion of Human Genome, %
Protein-coding genes			
Solitary genes	~1 kb	1	~1% (2.5%)
Duplicated or diverged genes (functional gene families and nonfunctional pseudogenes)	Variable	2-100	~2% (3.0%)
Tandemly repeated genes encoding rRNA, 5S rRNA, tRNA, and histones	~1 kb	10-100	~1%
Repetitious DNA			
Simple sequence DNA	~10-100 bp	Variable	~1%
Moderately repeated DNA (mobile DNA element)			
Transposons	~1 kb	10-100	~1%
Viral retrotransposons	~1 kb	10-100	~1%
Long interspersed elements (LINES; nonviral retrotransposons)	~1 kb	10-100	~1%
Short interspersed elements (SINES; nonviral retrotransposons)	~1 kb	10-100	~1%
Unclassified spacer DNA	Variable	~10	~2%

Major Classes of Eukaryotic DNA in Human Genome

Protein coding genes

Tandemly repeated genes

Repeated DNA

Unclassified spacer DNA

51 Duplicated Genes/Pseudo-Genes

Duplicated Genes

Encode closely related (homologous) proteins

Clustered together in genome

Formed by duplication of an ancestral gene followed by mutation

Duplicated Genes/Pseudo-Genes

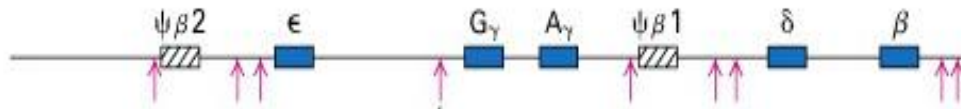
Pseudo-Genes

Nonfunctional copies of genes

Formed by duplication of ancestral gene, or reverse transcription and integration.

Five functional genes and two pseudo-genes

Human β -globin gene cluster (chromosome 11)



Not expressed due to mutations that produce

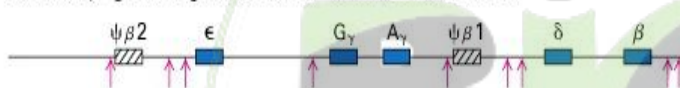
A stop codon

Prevent mRNA processing

Due to lack of regulatory sequences

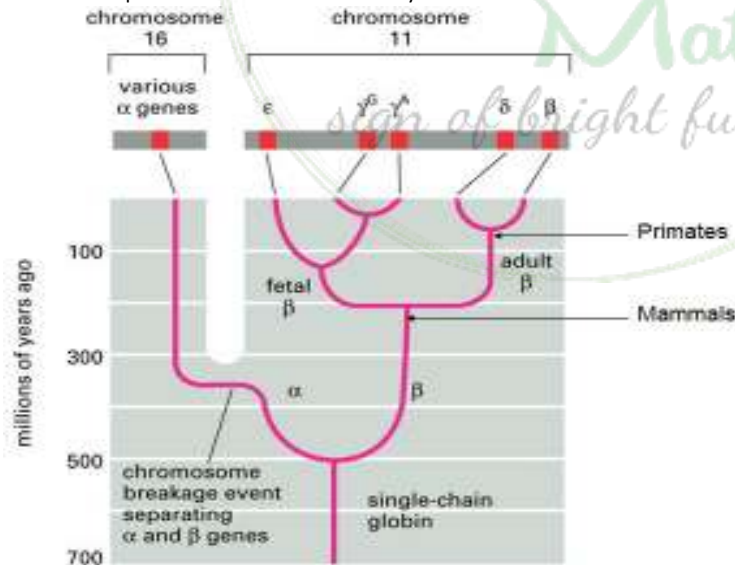
Five functional genes and two pseudo-genes

Human β -globin gene cluster (chromosome 11)



Duplicated Genes/Pseudo-Genes

Genes duplication: Globin family



Ancestral globin gene was duplicated ~ 500 million years ago.

Mutations in both genes to differentiate them - α and β present in all higher vertebrates

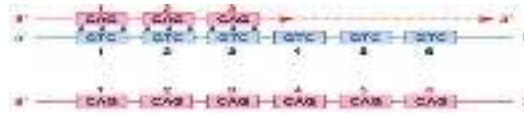
Further gene duplications produced alternative forms in mammals and in primates

DNA Duplications:

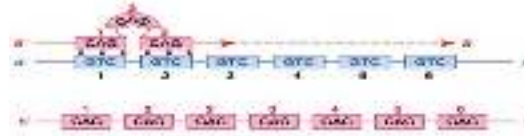
Two phenomenon are common in DNA duplications:

Slipped strand mispairing

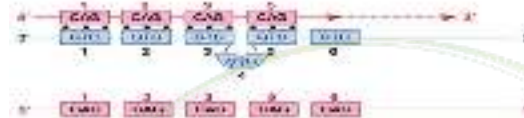
Normal



Backward



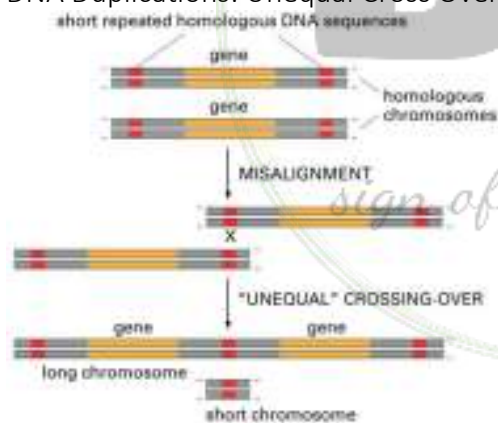
Forward
slippage



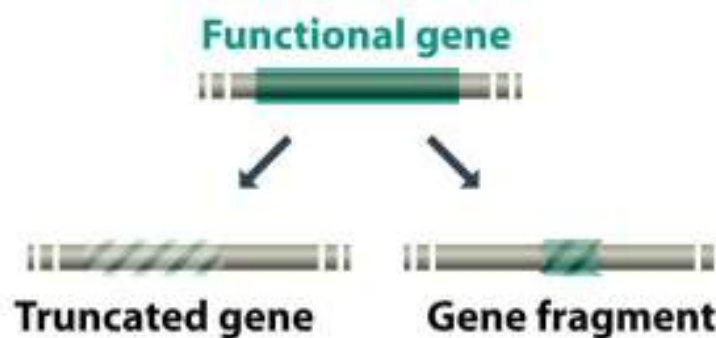
Unequal crossover during recombinations

DNA Duplications: Slipped strand mispairing

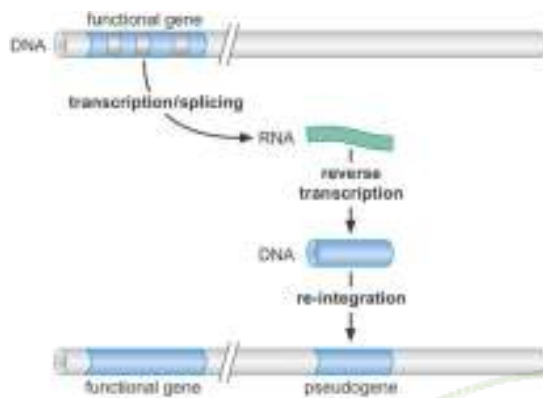
DNA Duplications: Unequal Cross Over



Functional Genes: Process of evolution produced truncated genes and gene fragments



Duplicated Genes/Pseudo-Genes



Duplicated Genes and Pseudo Genes

Clustered together in genome.

Formed by duplication of an ancestral gene

Nonfunctional copies of genes

52 Repetitive DNA

Whole genome duplication

The genomes of two distinct species can merge

An individual can acquire an extra copy of a chromosome

Chromosomes can fuse; e.g. human chromosome 2 derived from a fusion of two ancestral primate chromosomes

Chromosomal regions can be inverted or deleted

Segmental and other duplications can occur

Five main classes of Repetitive DNA

Interspersed repeats (RNA/DNA transposon-derived)

Approximately 45% of human genome (e.g. LINES, SINES, Alu)

Retrotransposons constitute over 40% of the human genome and consist of several millions of family members.

They play important roles in shaping the structure/evolution of the genome.

Processed Pseudogenes

These genes have a stop codon, frameshift mutation, or loss of promoter activity and do not encode a functional protein.

They commonly arise from retrotransposition, or following gene duplication and subsequent gene loss.

Repetitive DNA

Simple Sequence Repeats

Simple sequence repeats

Microsatellites (1-12 bp)

Minisatellites (12-500 bp)

Segmental Duplications

Segmental duplications blocks of about 1 kb to 300 kb that are copied within or between chromosomes

5% of human genome.

Blocks of Tandem Repeats

Blocks of tandem repeats includes telomeric and centromeric repeats and can span millions bp

Often species specific.

Repeats in Mouse genome

Type	No. of Repeats	Size	Percent of genome
Highly repetitive	> 1 Million	< 10 bp	10 %
Moderately repetitive	> 1000	~ 150 - ~300 bp	20 %

Conclusion

Main classes of Repetitive DNA

Simple sequence of DNA

DNA transposons

LTR retrotransposons

Non LTR retrotransposons like LINE, SINE

53 Mobile DNA

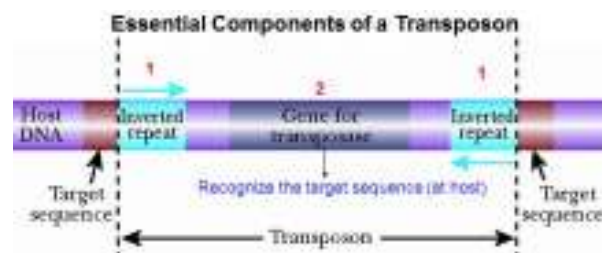
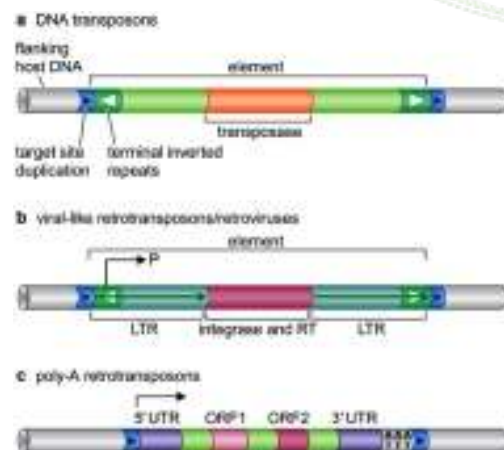
Move within genomes

Most of moderately repeated DNA sequences found throughout higher eukaryotic genomes

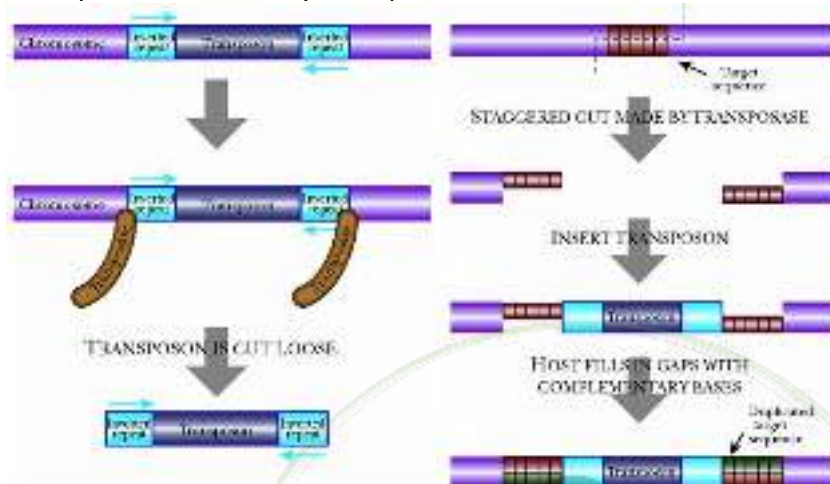
Most mobile DNA -transposable elements

Includes DNA-based transposons and retro-transposons.

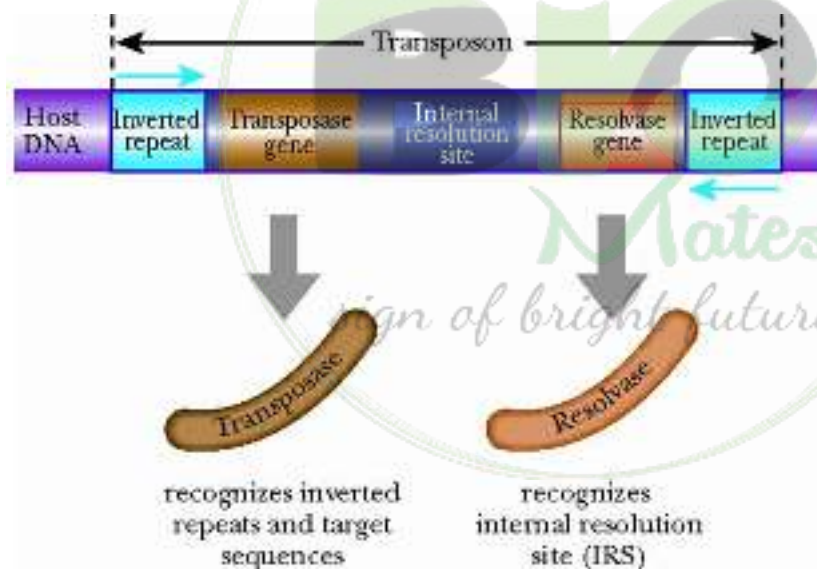
Some encode enzymes that catalyze movement.



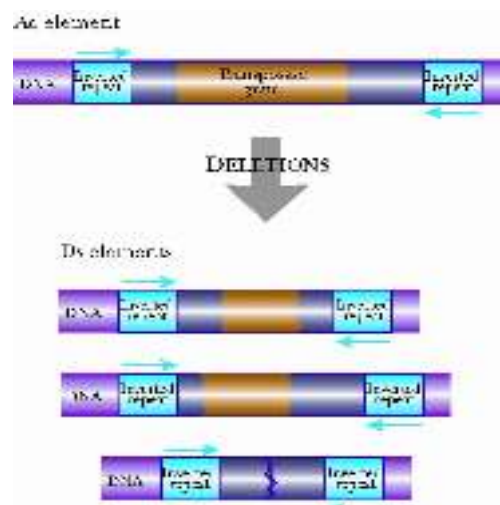
Transposon is cut off by transposase and inserted on another site



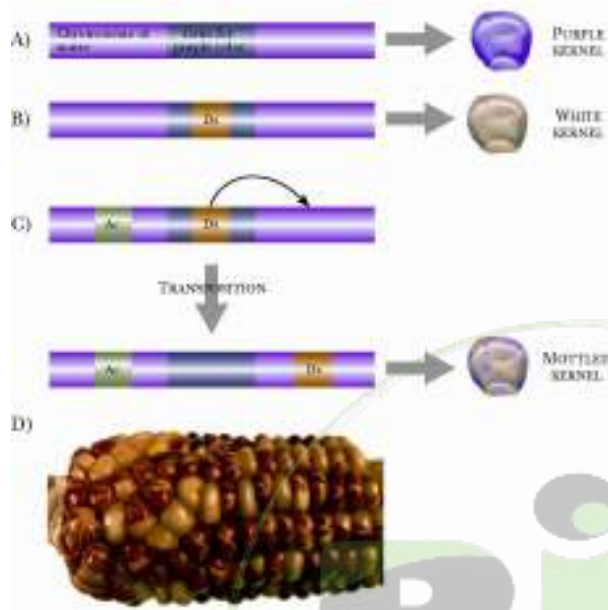
Components of a Complex Transposon



Ac and Ds Elements of Maize – Higher Life Form



Movement of Ds Elements Gives Mottled Corn



Mobile DNA Transpositions

Transposition is movement of mobile DNA

Involves copying of mobile DNA element and insertion into new site in genome

Molecular parasite: "selfish DNA"

Mobile DNA - Retrotransposition

Moving in the form of RNA by element coding for reverse transcriptase

- LINES
- SINEs
- retrovirus-like elements (long terminal repeat)

Mobile DNA – Interspersed Repeats in Human Genome

Element	Transposon	Structure	Length	Copy number	Fraction of genome
LINES	Autonomous	ORF1 ORF2 (pol)	7-8 kb	30,000-40,000	21%
SINEs	Nonautonomous	A B	100-300 bp	1,500,000	13%
Retrovirus-like elements	Autonomous	gag act (env)	8-11 kb	450,000	9%
	Nonautonomous	(gag)	1.3-2 kb		
DNA transposons	Autonomous	transposase	2-3 kb	300-500	3%
	Nonautonomous		80-2000 bp		

Mobile DNA – Fuel for Evolution

Probably have significant effect on evolution by facilitating gene duplication, which provides the fuel for evolution, and exon shuffling

54 Movement of Transposons and Retro-transposons

Transposons/Retrotransposons-Movement

Movement of Transposons

Eukaryotic transposable elements - two types

Transposons, which move by means of a DNA intermediate

Retrotransposons, which move by means of an RNA intermediate

Transposons - major content of eukaryotic genomes

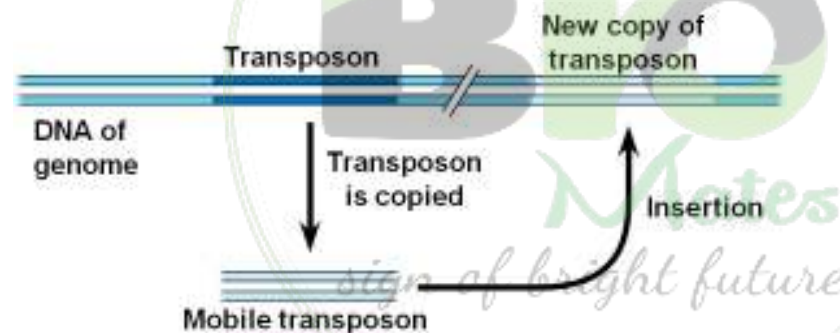
~50% of Human/Mouse

~75% of the maize genome

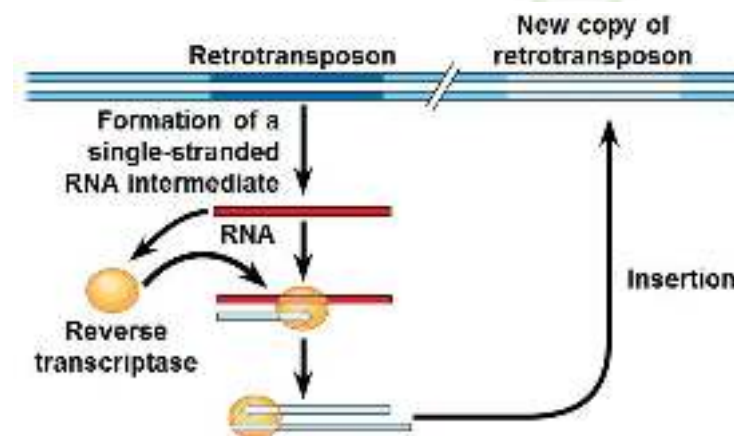
~85% of the barley genome

~98% of the iris genome

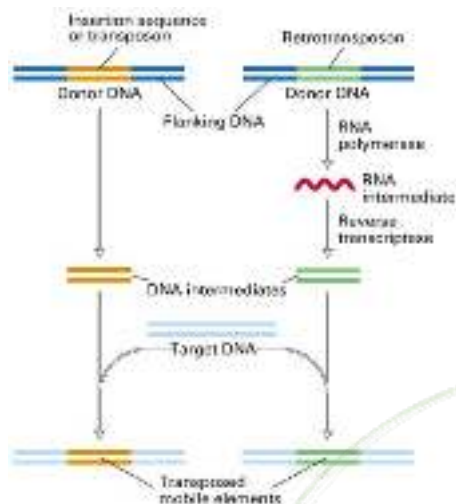
Movement of Transposons



Movement of Retrotransposons



Transposons moves using DNA intermediate



Conclusion

Eukaryotic transposable elements - two types

Transposons, which move by means of a DNA intermediate

Retrotransposons, which move by means of an RNA intermediate

55 Transposons in Prokaryotes and Eukaryotes

Transposons in Prokaryotes

Two types of transposons in prokaryotes i.e.

- Insertion Sequence (IS)
- Transposons (Tn)
- Bacteriophage Mu

IS Elements

Prokaryotic IS elements range in size from 768 bp to over 5 kb.

IS1 is 768 bp long, and present in 4–19 copies on the E. coli chromosome.

IS2 has 0–12 copies on the chromosome, and 1 copy on the F plasmid

IS10 is found in R plasmids

The ends of all sequences (IS elements) show inverted terminal repeats (IRs) of 9–41 bp

(e.g., IS1 has 23 bp of nearly identical sequence).

Transposons in Prokaryotes

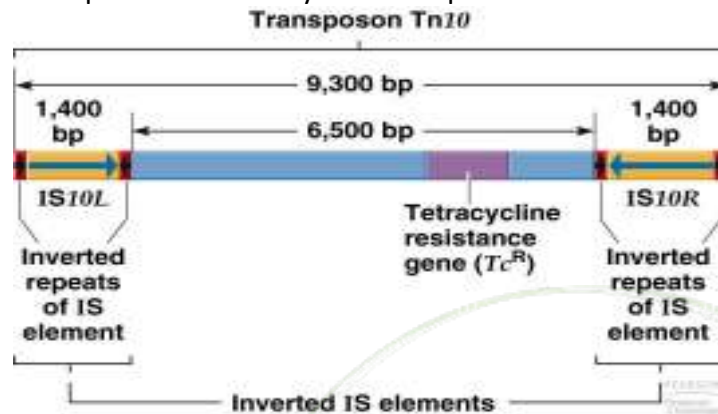
Transposons are similar to IS elements, but carry additional genes, and have a more complex structure.

Two types of prokaryotic transposons:

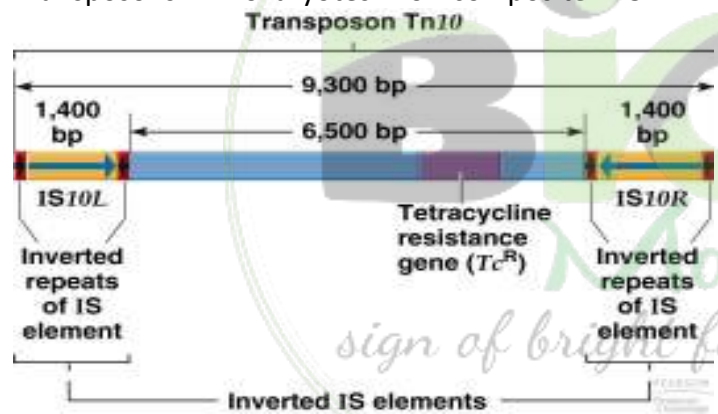
Composite transposons carry genes (e.g., antibiotic resistance) flanked on both sides by IS elements (IS modules). Example Tn10

Non-composite transposons also carry genes (e.g., drug resistance) but do not terminate with IS elements.

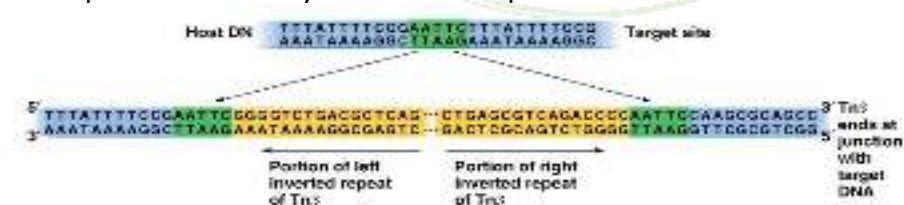
Transposons in Prokaryotes: Composite Tn10



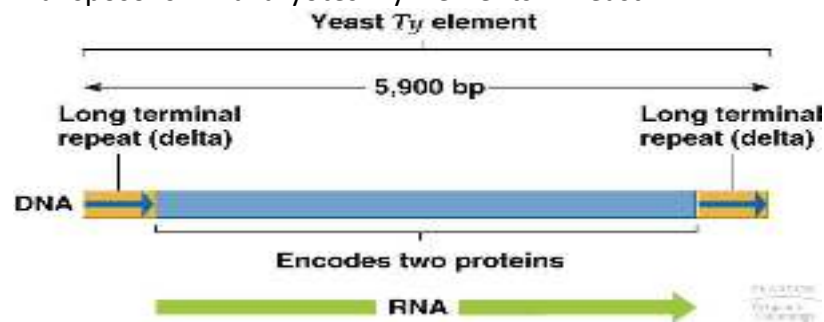
Transposons in Prokaryotes: Non-composite Tn3



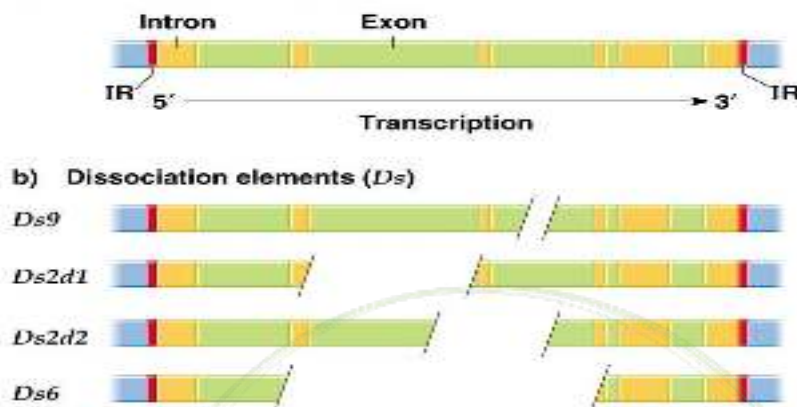
Transposons in Prokaryotes: Non-composite Tn3



Transposons in Eukaryotes: Ty Elements in Yeast



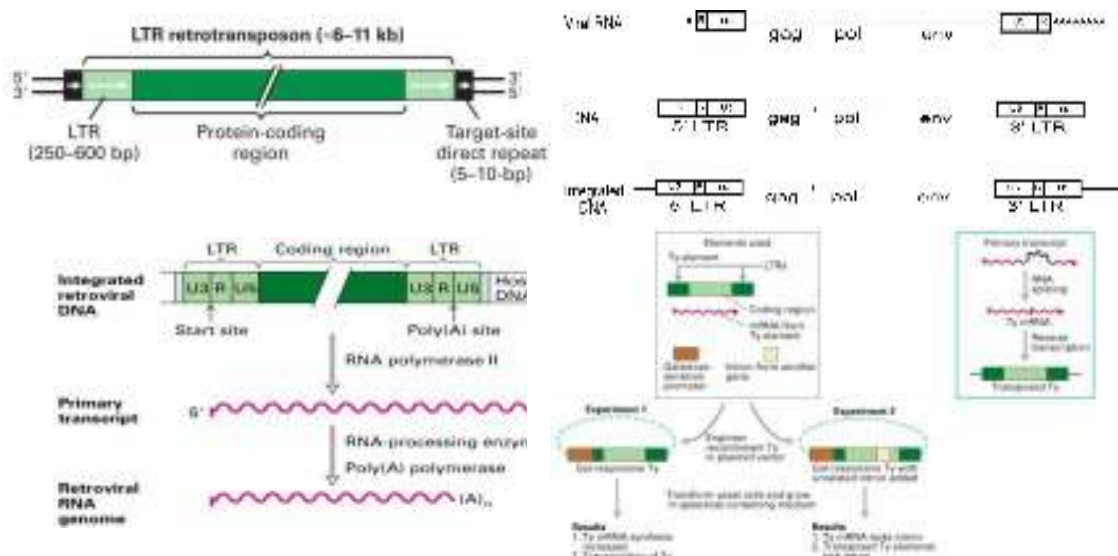
a) Activator element (A_c)



Ds elements vary in length and sequence, but all have the same terminal IRs as Ac, and many are deleted or rearranged versions of Ac

Transposons are found in both Prokaryotes and Eukaryotes.

Flank viral retrotransposons and retroviruses



Contain regulatory sequences like transcription start site and poly (A) site

Conclusion

Eukaryotic retrotransposons fall into two major groups:

LTR retrotransposons

Non-LTR retrotransposons

57 Long Interspersed Nuclear Elements and

Short Interspersed Nuclear Elements

LINEs and SINEs

Non-LTR retrotransposons consist of two sub-types;

Long Interspersed Nuclear Elements (LINEs)

Short Interspersed Nuclear Elements (SINEs)

High copy numbers, in the plants species

Widespread in eukaryotic genomes

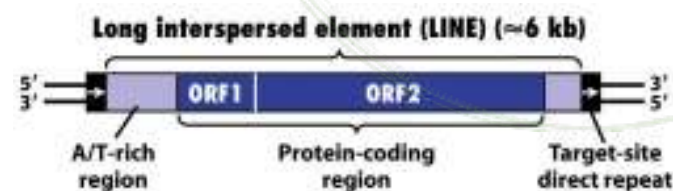
LINEs possess two ORFs, which encode all the functions needed for retrotransposition

LINEs

Functions of ORFs include

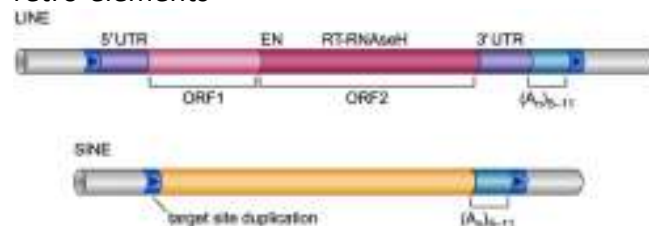
Reverse transcriptase

Endonuclease activities, in addition to a nucleic acid-binding property needed to form a ribonucleoprotein



SINEs

SINEs, on the other hand, use the LINE machinery and function as non-autonomous retro-elements



LINEs

Several subgroups, such as L1, L2 and L3.

Human coding L1 begin with an untranslated region (UTR) that includes an RNA polymerase II promoter, two non-overlapping open reading frames (ORF1 and ORF2), and ends with another UTR.

ORF1 encodes an RNA binding protein and ORF2 encodes a protein having an endonuclease

The 5' UTR contains the promoter sequence, while the 3' UTR contains a polyadenylation signal (AATAAA) and a poly-A tail

Human genome contains about 500,000 LINEs

LINEs and SINEs

Short Interspersed Elements

Short Interspersed Elements

~300 base pairs

Alu

~ 11-13% of human genome

Short Interspersed Elements

Short DNA sequences ~ 300 bases that represent reverse-transcribed RNA molecules

SINEs do not encode a functional reverse transcriptase protein and rely on other mobile elements for transposition

In some cases they may have their own endonuclease that will allow them to cleave their way into the genome

Short Interspersed Elements

The most common SINEs in primates are called Alu sequences.

Alu elements are approximately 350 base pairs long, do not contain any coding sequences, and can be recognized by the restriction enzyme AluI

SINEs make up ~ 11-13% of human genome

Conclusion

LINEs

6-7 kb long

SINES

~ 300bp long

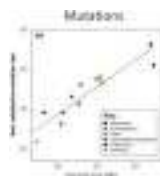
Example: Alu ~ 350 bp long

58 Genetic Variations

Variation between individuals of a population (within species) can be due to differences in the nucleotide sequence

How genetic variations emerge ?

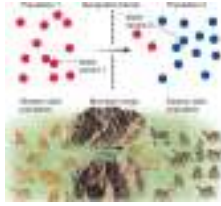
Genetic Variations can be due to the Mutations



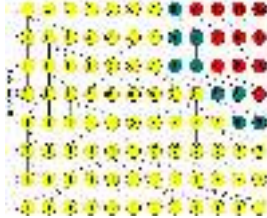
Genetic Variations: Recombinations



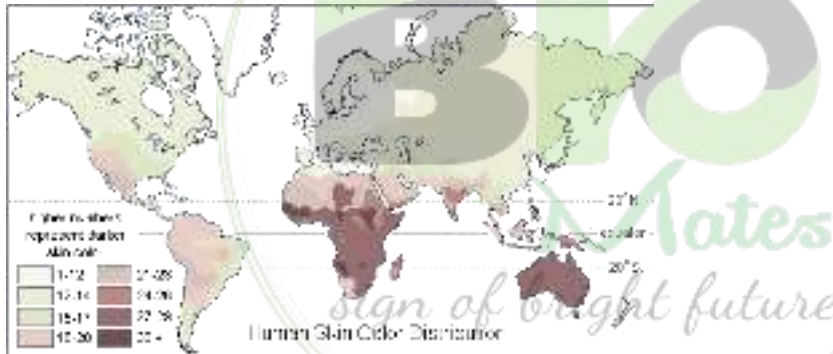
Genetic Variations: Genes Flow



Genetic Variations: Genetic Drift: Change in the frequency of a gene variant (allele) in a population due to random sampling of organisms

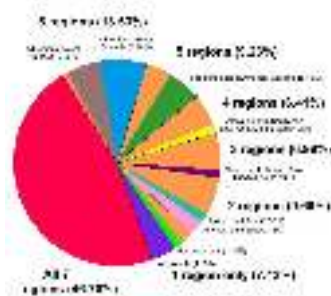


Genetic Variations: Humans Skin Color



Genetic Variations: Yellow Brown Hair Australian natives

Genetic Variations: Microsatellites



Genetic Variations - Humans

Two genomes are roughly 99.9% identical to each other.

If human genome is 3.0 billion bp then, there are 3.0 million differences between any two genomes.

59 Basic techniques

The solutions: basic techniques

The fragments of DNA can be inserted to a suitable replicons-cloning vehicles

Plasmids and bacteriophages are most suitable vectors

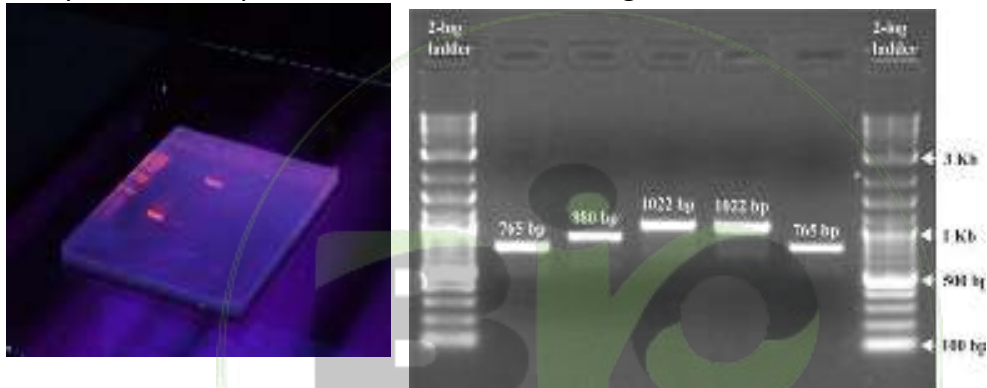
Isolation of genomic DNA from E. coli

Harvesting of cell pellet by centrifugation

Cell lyses by SDS and proteinase K
DNA extraction with phenol and chloroform
Precipitation with isoprenol

60 Agarose gel electrophoresis

Agarose is convenient for separating DNA fragments ranging in size from a few hundred base pairs to about 20 kb.
Polyacrlamide gel electrophoresis
Polyacrlamide is preferred for smaller DNA fragments



Nucleic acid blotting

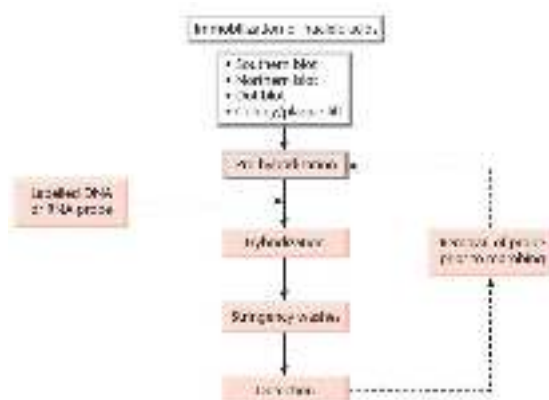
Blotting describes the immobilization of sample nucleic acids or proteins on to a solid support, generally nylon or nitrocellulose membranes
Southern blotting
Northern blotting
Western blotting

Nucleic acid Hybridization

The hybridization of nucleic acids on membranes is a widely used technique in gene manipulation

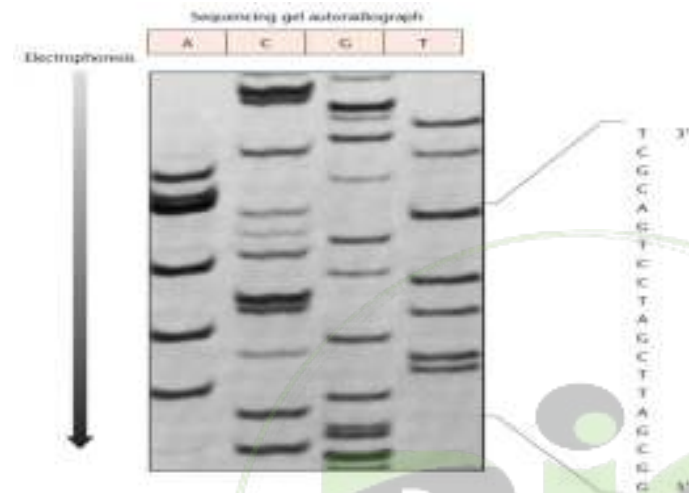
Stringency control

Stringency can be regarded as the specificity with which a particular target sequence is detected by hybridization probe
Probe is a fragment of DNA or RNA which is used to detect the sequences of target DNA



Autoradiography

A technique using X-ray film to visualize fragments of molecules that have been radioactively labeled. For example, it can be used to analyze the length and number of DNA fragments separated by gel electrophoresis.



Southern blotting

A Southern blot is a laboratory method used to detect specific DNA fragment from a mixture of DNA molecules.

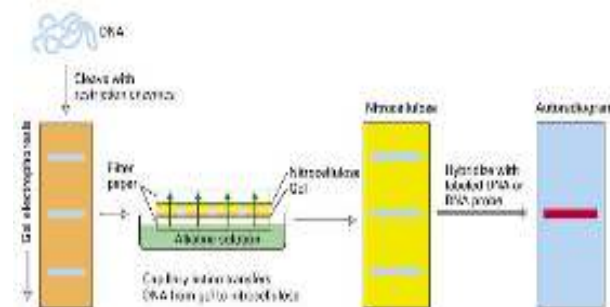
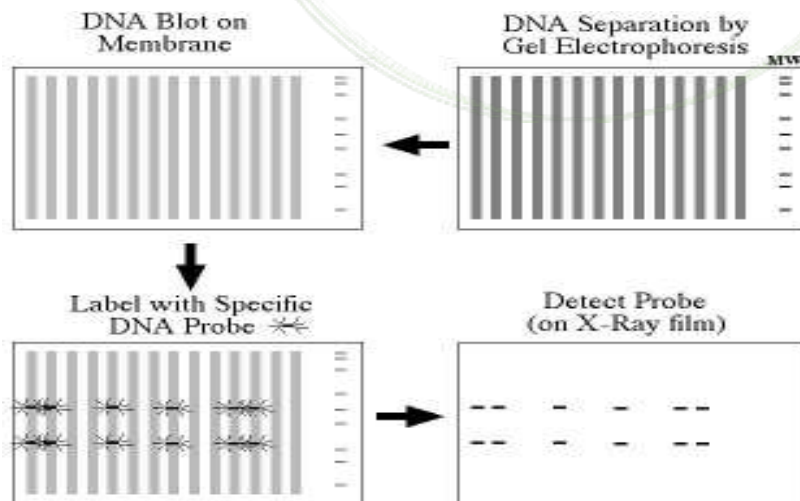
The technique was named after its inventor, Edward Southern (1975).

For efficient blotting, gel pretreatment is important.

After transfer, the nucleic acid needs fixation.

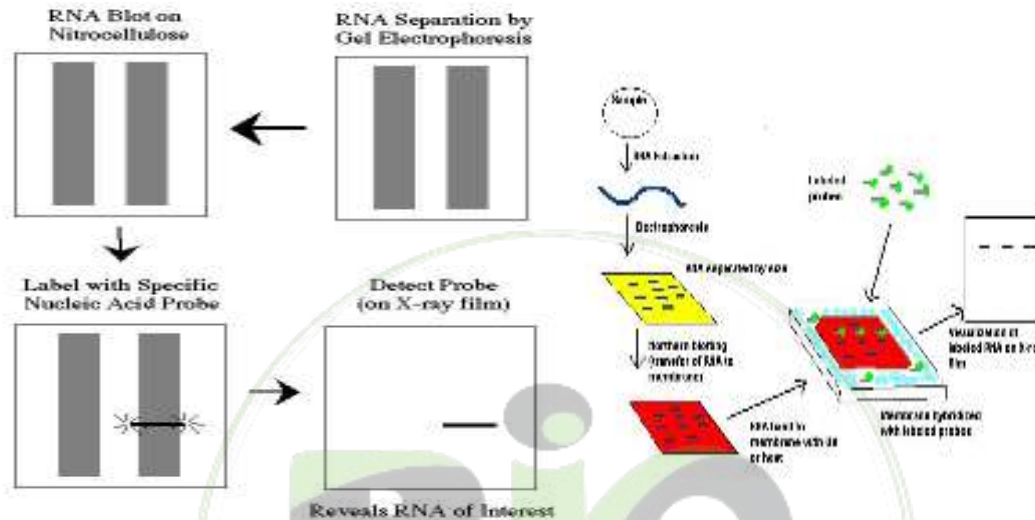
After fixation, the membrane is placed in a solution of labelled probe.

After hybridization reaction, membrane is washed and hybridization is detected by autoradiography.



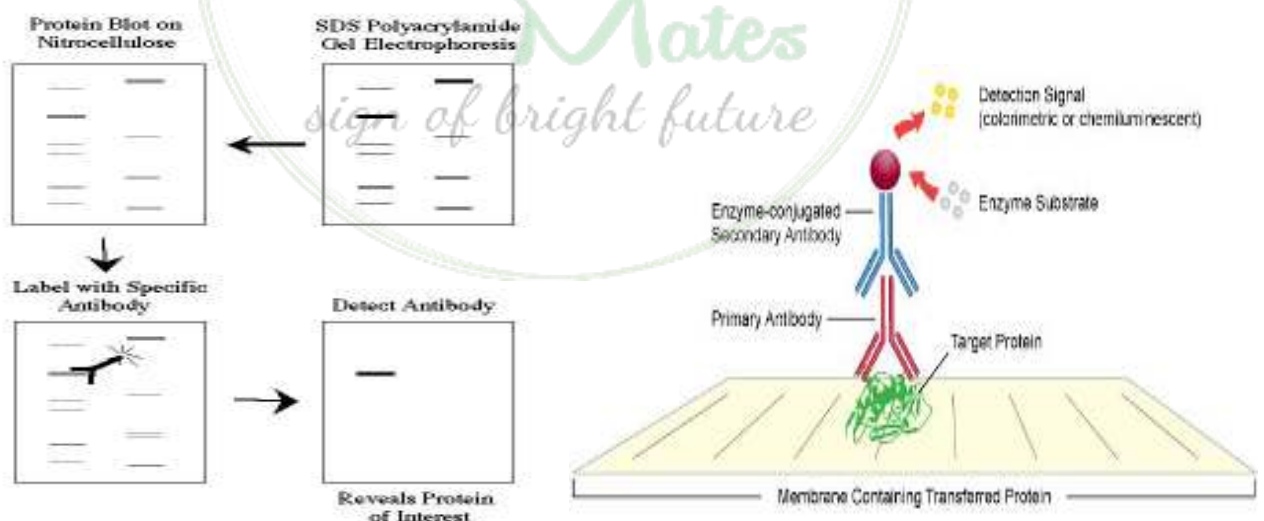
Northern blotting

A northern blot is a method used to detect specific RNA molecules among a mixture of RNA. Northern blotting can be used to analyze a sample of RNA from a particular tissue or cell type in order to measure the RNA expression of particular genes.



Western blotting

It involves the transfer of electrophoresed protein bands from a gel on to a membrane. The bound proteins are then available for analysis. Most commonly, antibodies are used to detect specific antigens.



67 Transformation of E. coli

Transformation is the process of getting the recombinant vector from a reaction mixture or vector solution into E. coli cells.

To enable the cells to take up circular vector DNA they have to be made competent. Electroporation

A rapid and simple technique for introducing cloned genes into a wide variety of microbial, plant and animal cell, including E. coli is electroporation.

In electroporation, electric field is applied to cells in order to increase permeability of the cell membrane to take DNA.

Transformation with other organisms

E. coli often remains the host of choice for cloning.

Bacillus subtilis is a particularly attractive alternative cloning host.

Animal cells, protoplasts of yeast and plant are susceptible to transformation by liposomes.

Polymerase chain reaction (PCR)

The impact of PCR upon molecular biology has been profound

The reaction is easily performed and leads to the amplification of specific DNA by an enormous factor

Cycle number	Number of double-stranded target molecules
1	0
2	0
3	2
4	4
5	8
6	16
7	32
8	64
9	128
10	256
11	512
12	1024
13	2048
14	4096
15	8192
16	16,384
17	32,768
18	65,536
19	131,072
20	262,144
21	524,288
22	1,048,576
23	2,097,152
24	4,194,304
25	8,388,608
26	16,777,216
27	33,554,432
28	67,108,864
29	134,217,728
30	268,435,456

Cutting DNA molecules

Before 1970 there was no method of cleaving DNA at discrete points

Mechanical shearing was used for DNA fragmentation

During 1960s phage biologist elucidated the phenomenon of restriction and modification

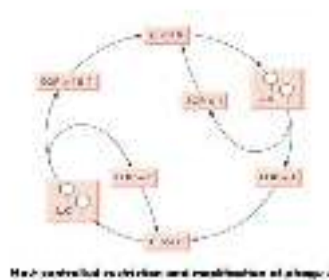
Host-controlled restriction/ modification

Restriction system allow bacteria to monitor the origin of incoming DNA

When the incoming DNA is a bacteriophage genome, the effect is to reduce the efficiency of plating i.e. to reduce the number of plaques in plating test

The phenomenon of restriction and modification were well illustrated and studied by the behavior of phage λ on two E. coli strains

E. coli C and E. coli K



Restrictive host must, of course protects its own DNA from restriction endonucleases by modification

Modification involves methylation of certain bases that constitute recognition sites

Types of restriction and modification (R-M) system

At least four R-M systems are known

Type I

Type II

Type III
Type IIs

Type I

Type I systems were the first to be characterized from *E. coli* K12

The active enzyme consists of two restriction subunit, two modification subunit and one recognition subunit

Type I systems are of little value for gene manipulation

Type II

Most of the useful R-M system is Type II

Type II enzymes recognize a defined sequence and cut within it

Type III

Type III enzymes have symmetrical recognition sequences but otherwise resemble type I systems and are of little value

Type IIs

Type IIs systems have similar cofactors and structure to type II but restriction occurs at a distance from recognition site that limits their usefulness

Cutting DNA molecules

Nomenclature

A suitable system was proposed by Smith and Nathans (1973)

The species name of host organisms is identified by the first letter of genus and first two letters of specific epithet

E. coli = Eco

H. influenzae = Hin

Strain identification is written as EcoK

In case, host strain has several restriction and modification systems, these are identified by roman numerals, for example, in case of *H. influenzae*

HindI, HindII, HindIII etc

All restriction enzymes have general name endonuclease R and modification-methylase M followed by the system name, for example, in case of *H. influenzae*

R. HindIII or M. HindIII

Examples of restriction endonuclease nomenclature

Enzyme	Enzyme source	Recognition sequence
<i>SmaI</i>	<i>Serratia marcescens</i> , 1st enzyme	CCCGGG
<i>HaeIII</i>	<i>Haemophilus aegyptius</i> , 3rd enzyme	GGCC
<i>HindII</i>	<i>H. influenzae</i> , strain d, 2nd enzyme	GTPyPuAC
<i>HindIII</i>	<i>H. influenzae</i> , strain d, 3rd enzyme	AAGCTT
<i>HamHI</i>	<i>Bacillus amyloliquefaciens</i> , strain H, 1st enzyme	GGATCC

Target sites

Type II endonucleases recognize and cleave DNA within particular sequences of four to eight nucleotides which have a twofold axis of rotational symmetry i.e. referred as palindromes

5'-GAATTC-3'

5'-CTTAAG-3'

5'-G A A T T C-3'

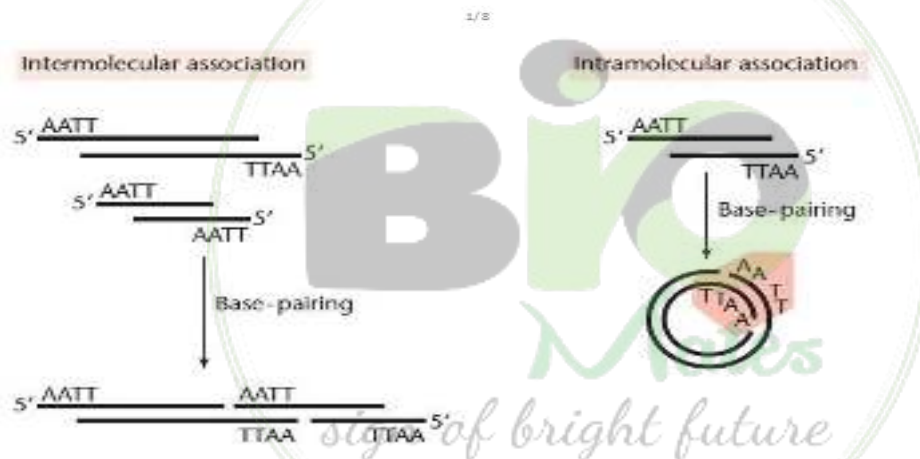
3'-C T T A A G-5'

5'-G / A A* T T C-3'

3'-C T T A* A / G-5'

5'-G 5'- A A T T C-3'
3'-C T T A A-5' G-5'

Single stranded breaks by *EcoR1*



Number and size of restriction fragments

The number and size of fragments generated by restriction enzyme depend on the frequency of occurrence of the target site in the DNA to be cut

Four base recognition site occurs every 44 (256) bp

Six base recognition site occurs every 46 (4096) bp

Eight base recognition site occurs 48 (65,536) bp

Average fragment size (bp) produced by different enzymes

Enzyme	Target	Arabidopsis	E. Coli	Human
<i>Apal</i>	GGGCCC	25000	15000	2000
<i>BamHI</i>	GGATCC	6000	5000	5000
<i>SpeI</i>	ACTAGT	8000	60000	10000

Summary of restriction endonucleases

Type II restriction enzymes are heavily responsible for the current explosion in the field of gene manipulation in that they are essential in forming recombinant DNA molecules

77 Joining DNA molecules

DNA modifying enzymes

- Nucleases
- DNA Polymerase
- Reverse transcriptases
- DNA ligases

Nucleases

Nucleases or DNases are the enzymes that degrade DNA

Two broad classes of nucleases

- i). Exonucleases
- ii). Endonucleases

DNA Polymerase

An enzyme that catalyzes template-dependent synthesis of DNA

Reverse transcriptases

An RNA-directed DNA polymerase in retroviruses; capable of making DNA complementary to an RNA

DNA ligase

An enzyme that creates a phosphodiester bond between the 3' end of one DNA segment and the 5' of another

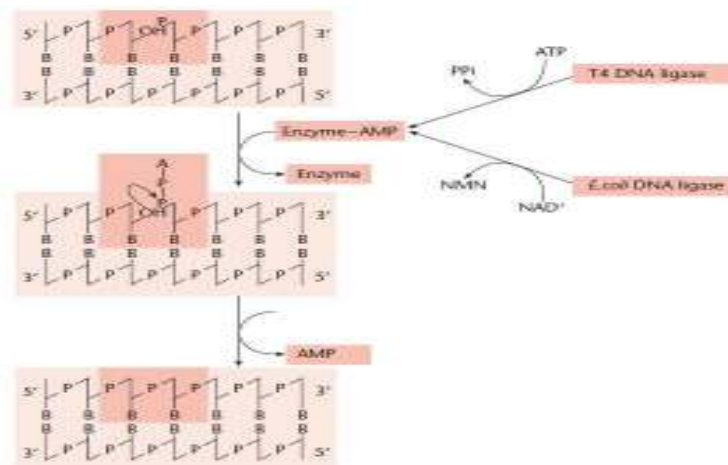
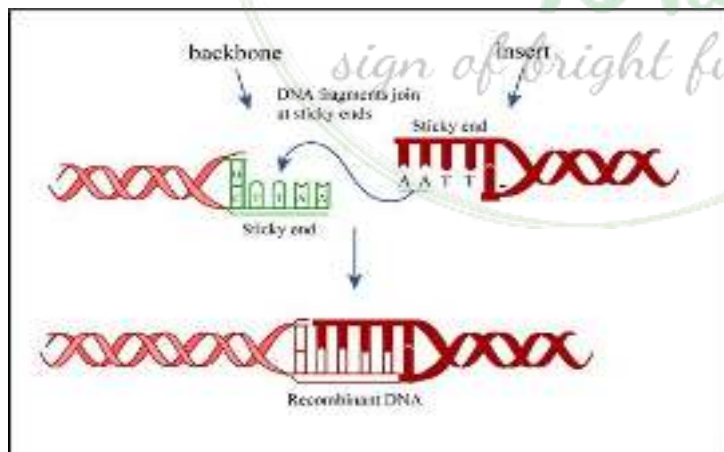
78 Methods of joining DNA fragments

Mainly three methods are used for joining DNA in vitro

- i). Joining covalently annealed cohesive ends by DNA ligase
- ii). Joining blunt-ended fragments by DNA ligase from phage T4 infected E. coli
- iii). The third utilizes the enzyme terminal deoxynucleotidyltransferase to synthesize homopolymeric 3' single stranded tails at the ends of fragments

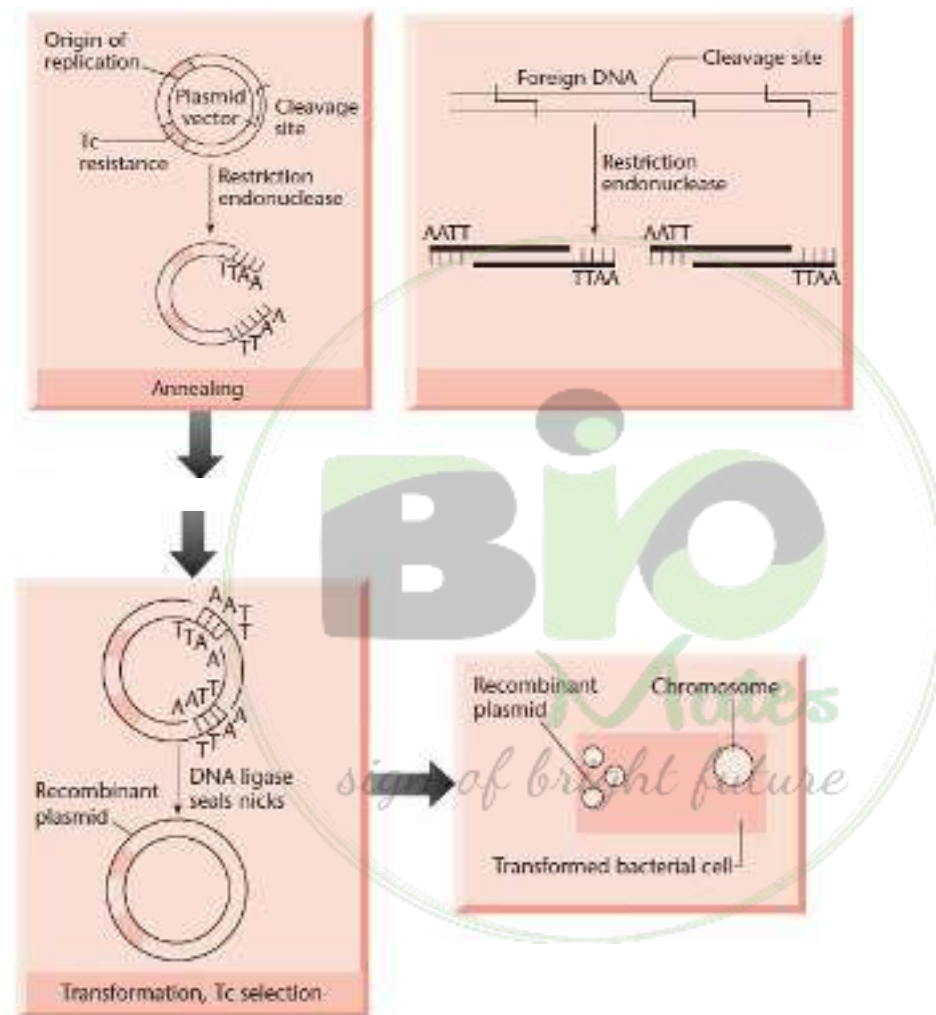
DNA ligase

E. coli and phage T4 encode enzyme, DNA ligase, which seals single stranded nicks between adjacent nucleotides in a duplex DNA chain



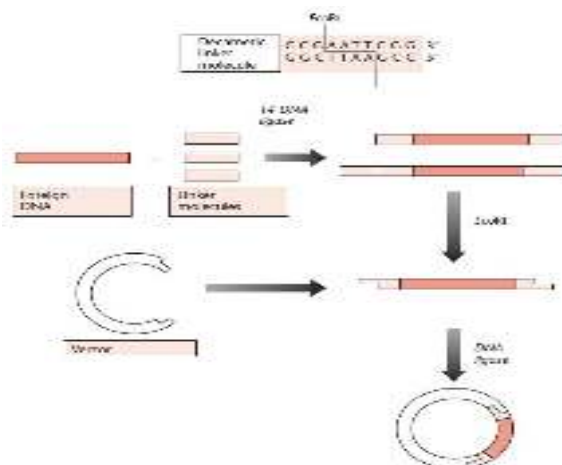
79 DNA ligase to create covalent recombinant DNA

DNA ligase can repair the nicks produced after the association of cohesive ends of DNA strands



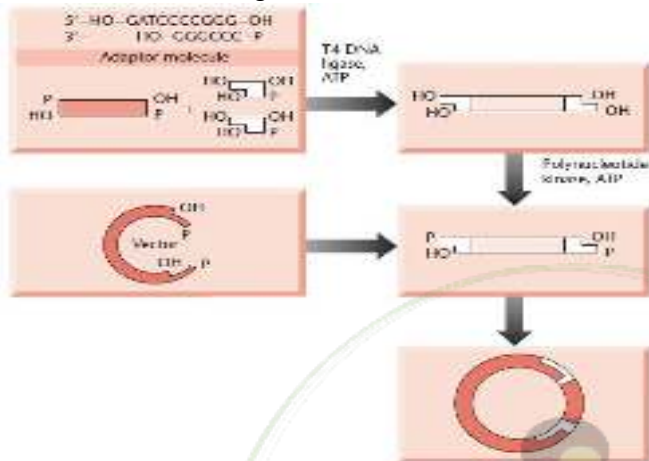
80 Blunt end ligation via linker molecules

Joining DNA fragments with cohesive ends by DNA ligase is a relatively efficient process. T4 DNA ligase has been used to join blunt-ended DNA molecules.



82 Adaptors

Chemically synthesized adaptor molecules which have a preformed cohesive end can be used to insert blunt ended foreign DNA into vector

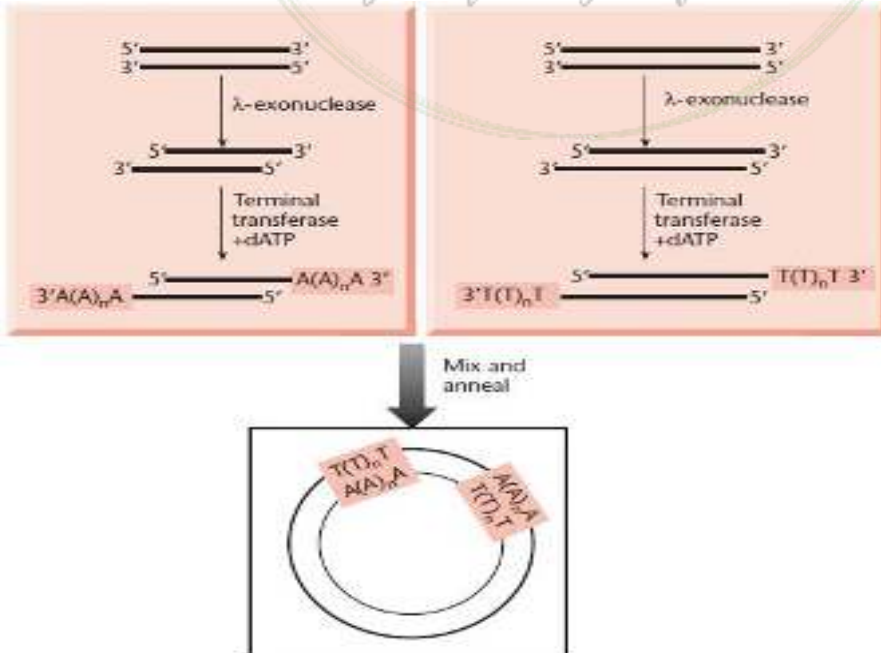


The main difference between linker and adaptor is that former having blunt ends while later have one cohesive end

83 Homopolymer tailing

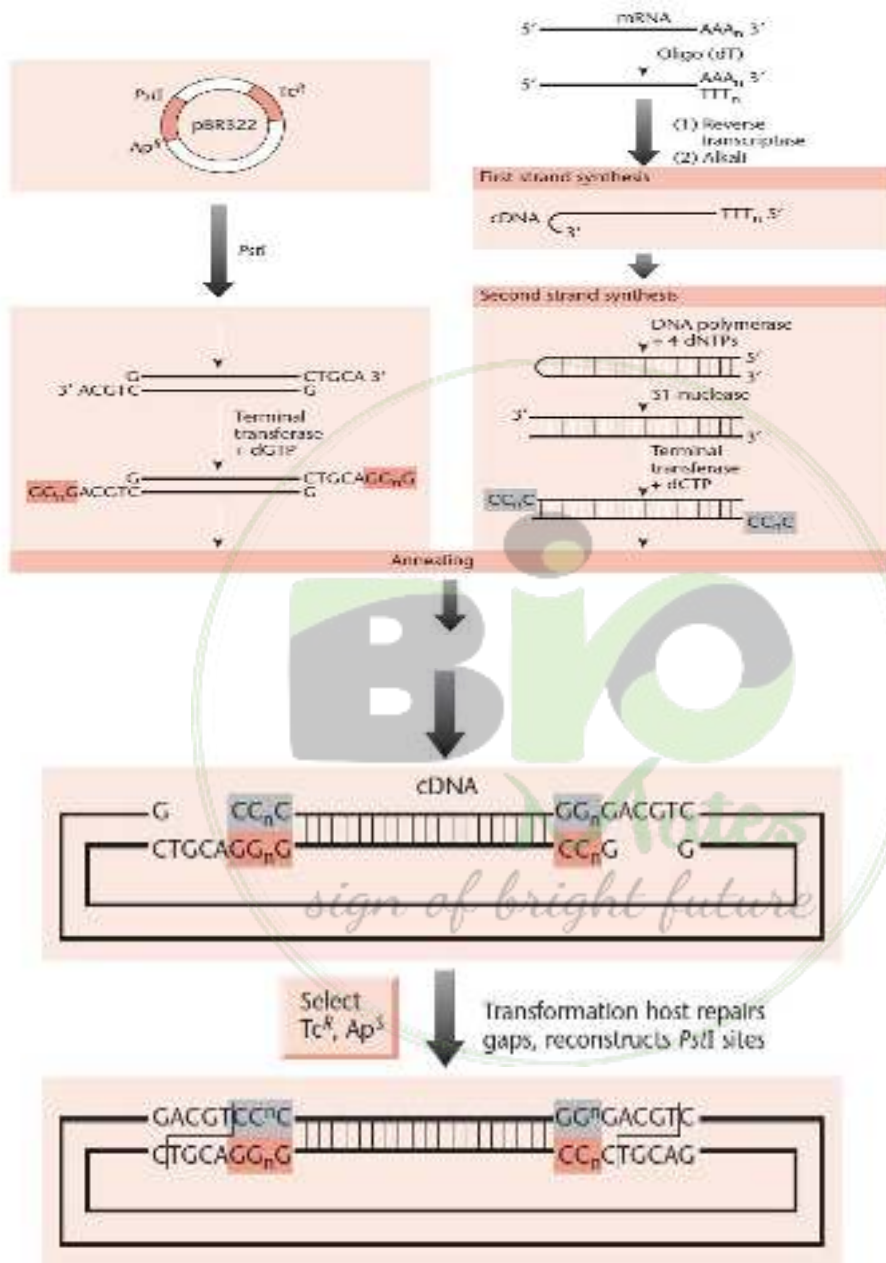
A general methods for joining DNA molecules makes use of the annealing of complementary homopolymer sequences by adding oligo(dA) sequences to 3' ends of one population of DNA molecules and oligo(dT) blocks to the 3' ends of another population

An enzyme purified from calf thymus, terminal deoxynucleotidyltransferase, provides the means by which the homopolymeric extensions at 3' OH termini can be synthesized, if presented with a single type of nucleotide



84 Cloning of cDNA by Homopolymer tailing

Eukaryotic mRNA can be cloned in vector after converting it to cDNA by using an enzyme reverse transcriptase



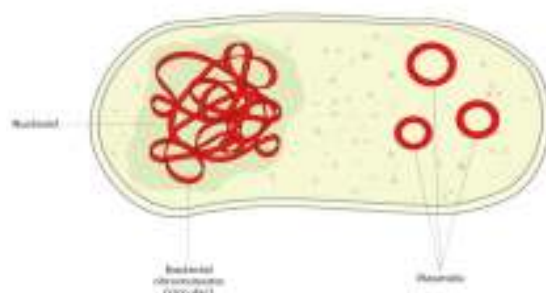
cutting and joining DNA molecules is basic step in gene manipulation

85 Plasmids

Plasmids are mostly circular DNA molecules that replicate separately from the host chromosome

Naturally occurring bacterial plasmids range in size from 5,000 to 400,000 bp.

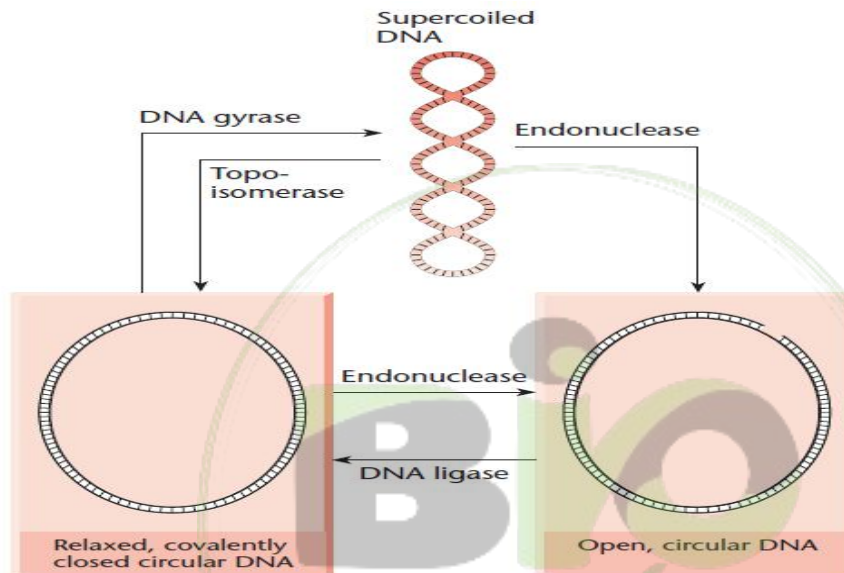
Plasmid help bacteria to survive stress



86 Interconversion of plasmid DNA

Most plasmids exist as double-stranded circular DNA molecules

If both strands of DNA are intact circles the molecules are describes as covalently closed circles or CCC DNA. Open circle or OC DNA have one strand intact

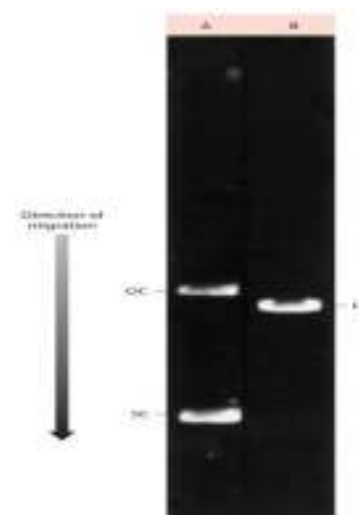


87 Effect of ethidium bromide on supercoiling of DNA

DNA bands can be visualized by soaking the gel in a solution of ethidium bromide

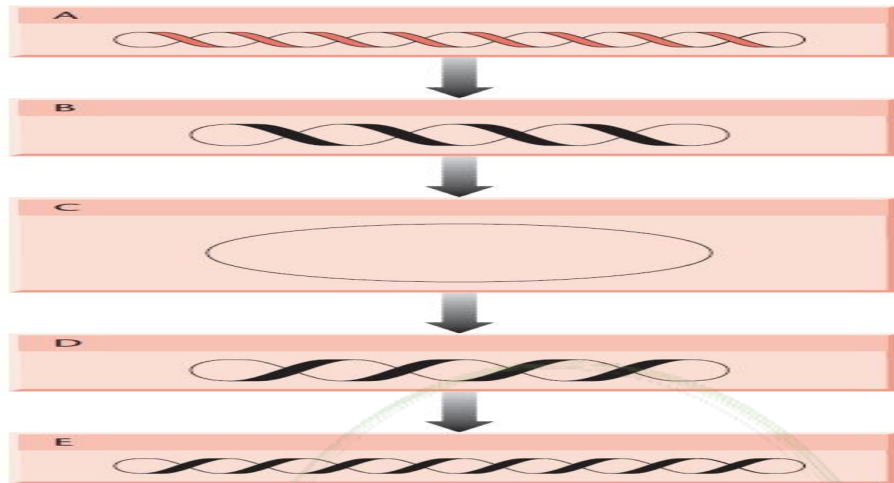
As the amount of intercalating agent i.e. ethidium bromide increases, the double helix untwists until the open form of circular DNA is produced

Electrophoresis of DNA in agarose gel



Visualization of gel under UV light after staining with ethidium bromide





Effect of ethidium bromide on supercoiling of DNA

Ethidium bromide may be a mutagen, a carcinogen, or a teratogen

88 Plasmid Biology

Phenotypic traits exhibited by plasmids

Plasmids are widely distributed throughout the prokaryotes, vary in size from less than 1×10^6 to greater than 200×10^6 daltons

Antibiotic resistance

Antibiotic production

Degradation of aromatic compounds

Haemolysin production

Sugar fermentation

Enterotoxin production

Heavy-metal resistance

Bacteriocin production

Induction of plant tumours

Hydrogen sulphide production

Host-controlled restriction and modification

89 Properties of Conjugative and non-conjugative plasmids

Plasmids can be categorized into conjugative or non-conjugative depending whether or not they carry a set of transfer genes, called tra genes, which promoted bacterial conjugation

Plas mid	Size (Mda)	Conj ugati ve	No. of plas mid copies	Phenotype
Col E1	4.2	No	10-15	ColE1 production
RSF 103 0	5.6	No	20-40	Ampicillin resistance
cloD F13	6	No	10	Cloacin production
R6k	25	Yes	13-38	Ampicillin and streptomycin resistance
F	62	Yes	1-2	-
RI	RI	Yes	3-6	Multiple drug resistance
Ent P 30 7	65	Yes	1-3	Enterotoxin production

Properties of some conjugative and non-conjugative plasmids of Gram-negative organisms

90 Host range of plasmids

Plasmids encode only a few of the proteins required for their own replication

Host range of plasmid is determined by ori region

Plasmids whose ori derived from Col E1 have restricted host range

Host range of plasmids

RP4 and RSF1010 and many plasmids from *Staphylococcus aureus* have broad host range

91 Partitioning and segregative stability of plasmids

The loss of plasmids due to defective partitioning is called segregative instability

Naturally occurring plasmids are stably maintained because they contain *par* regions

92 Incompatibility of plasmids

The inability of two different plasmids to coexist in the same cell in the absence of selection pressure

93 Desirable properties of plasmid cloning vehicles

An ideal cloning vehicle would have the following three properties

i). Low molecular weight

ii). Ability to confer readily selectable phenotypic traits on host cells

Desirable properties of plasmid cloning vehicles

iii). Single sites for number of restriction endonucleases

95 Natural plasmids as cloning vehicles

Plasmids which were not constructed in vitro for the sole purpose of cloning are called natural plasmids

Col E1 is a natural occurring plasmid

RSF2124 is a derivative of Col E1

pSC101 is natural plasmid from *Salmonella panama*

Plasmid	Size (Mda)	Sites of endonucleases	Selectable markers	Insertional inactivation
pSC101	5.8	<i>XhoI, EcoRI, PvuII, HincII, HpaI</i>	Tetracycline resistance	---
		<i>HindIII, BamHI, SalI</i>	—	Tetracycline resistance
Col E1	4.2	<i>EcoRI</i>	Immunity to colicin E1	Colicin E1 production
RSF2124	7.4	<i>EcoRI, BamHI</i>	Ampicillin resistance	Colicin E1 production

Cloning in pSC101,
ColiE1 and RSF2124

96 Use of pSC101 for cloning

Expression of *Staphylococcus* plasmid genes in *E. coli*

pSC101 contain

Tetracycline resistance

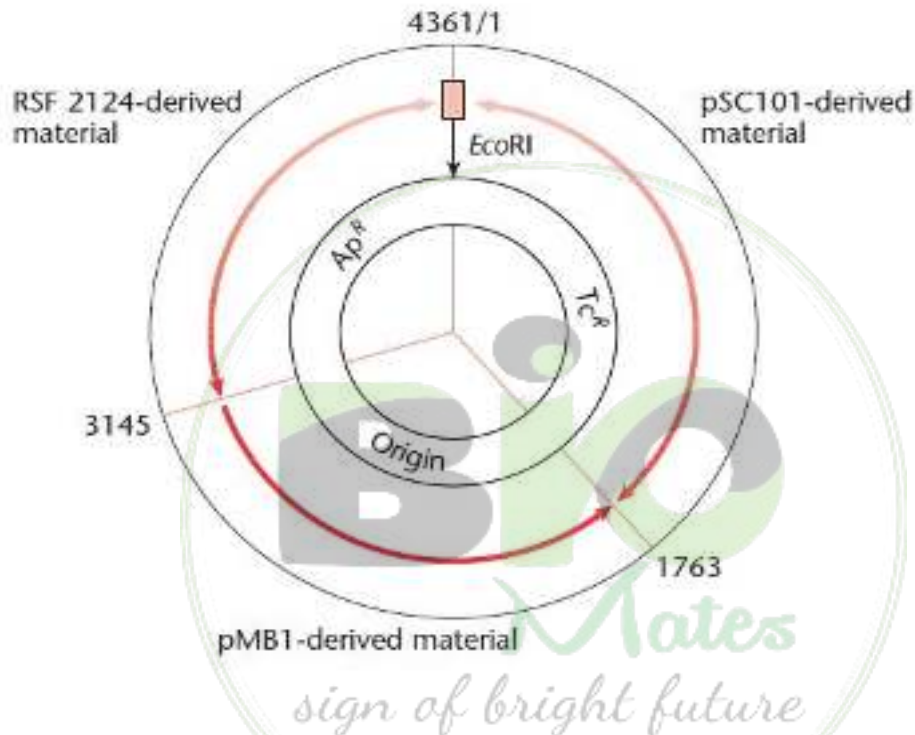
P1248 from

Staphylococcus

97 pBR322, a purpose-built cloning vehicle

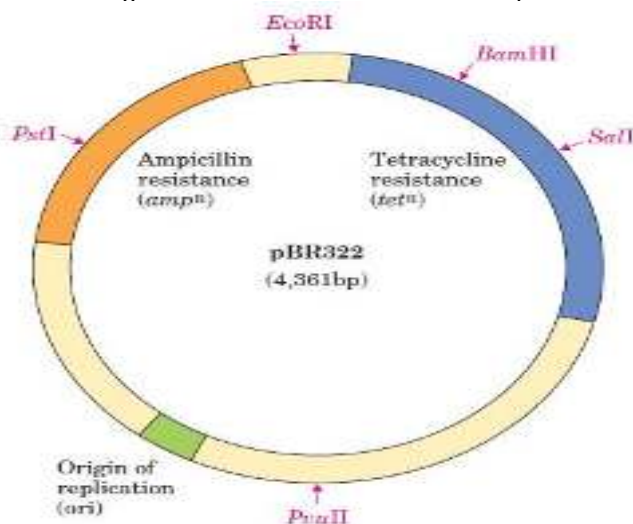
pBR322 is an example of in vitro constructed cloning vehicle

pBR322 contains the Ap^R and Tc^R genes of RSF2124 and pSC101, combined with replication element of pMB1



The origin of plasmid pBR322

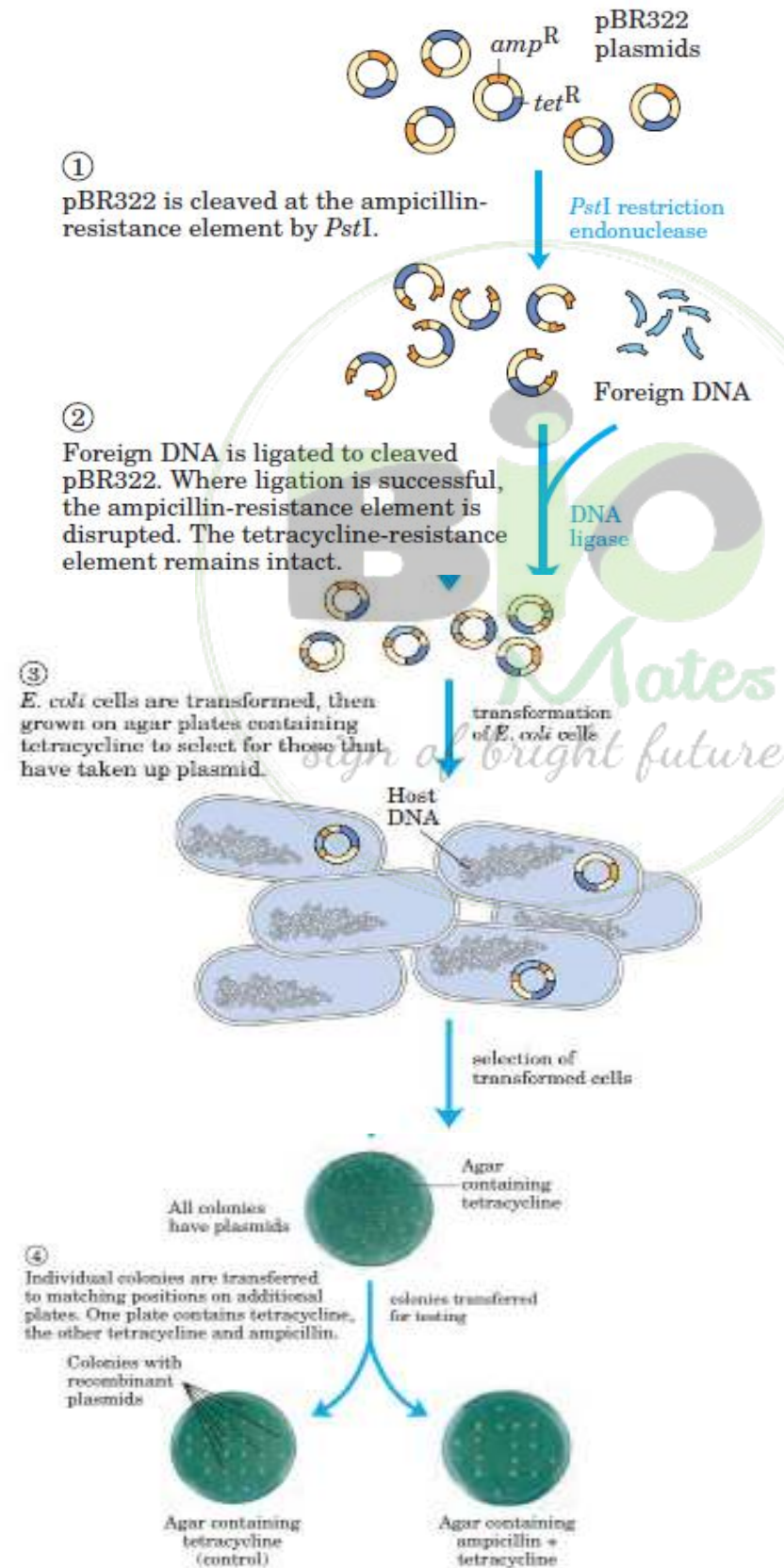
It has 40 target sites for different restriction enzyme



The constructed E. coli plasmid pBR322

98 Cloning with pBR322

Plasmid pBR322 has a number of characteristics that make it suitable as cloning vehicle



99 Improved vectors derived from pBR322

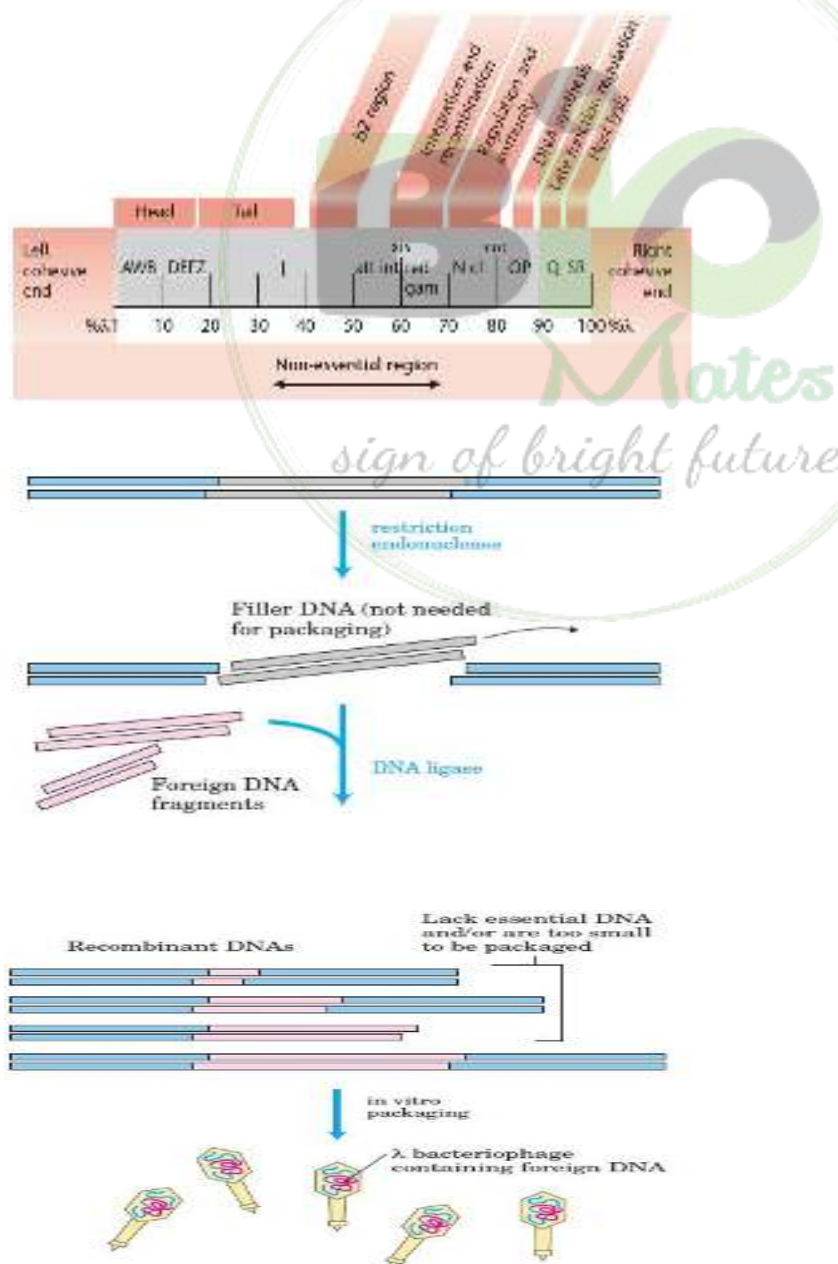
- Over the years numerous different derivatives of pBR322 have been constructed
- pBR325 encodes chloramphenicol resistance in addition to ampicillin and tetracycline
- pUC19 carrying sites for many different restriction endonucleases
- pAT153 another derivative of pBR322
- Loss of copy number control –runaway replication

100 Runaway plasmid vectors

Loss of copy number control –runaway replication

101 Bacteriophage λ as cloning vector

Bacteriophage λ has a very efficient mechanism for delivering its 48,502 bp of DNA into bacterium
About one-third of the λ genome is nonessential and can be replaced with foreign DNA

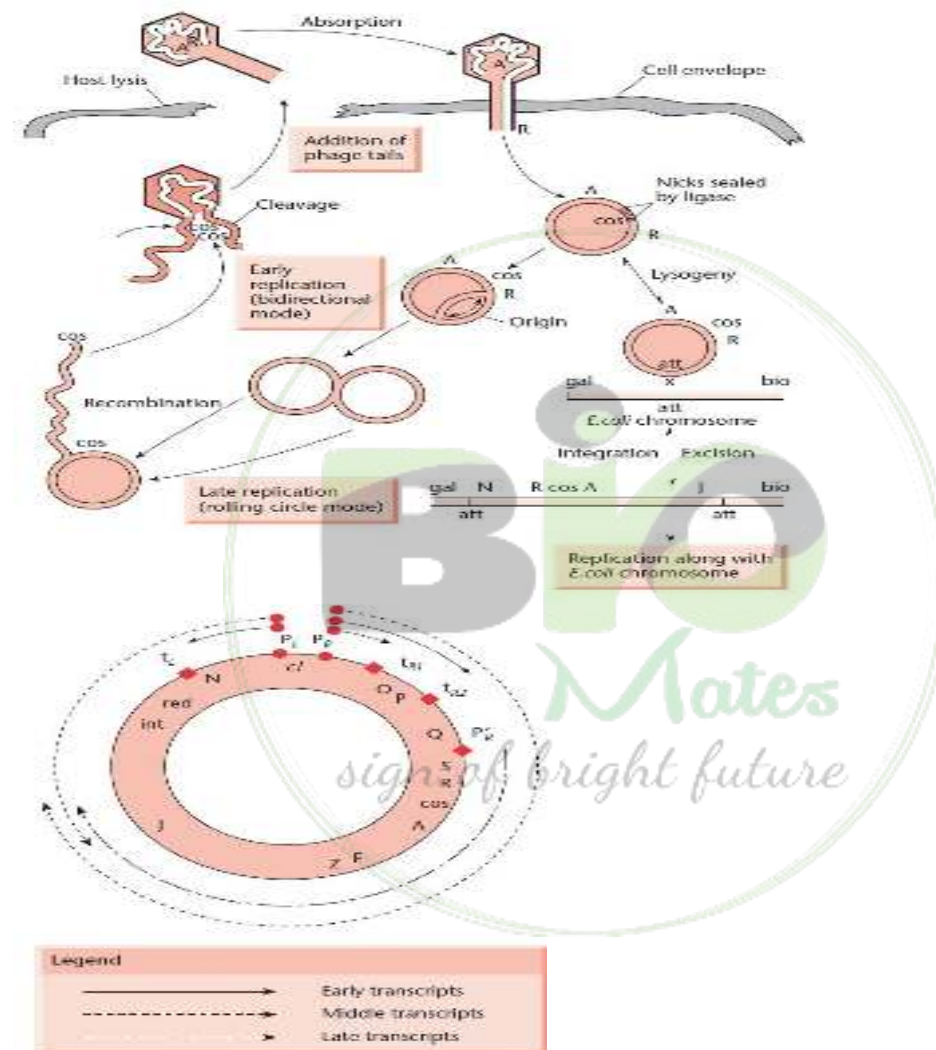


Bacteriophage λ permits the cloning of DNA fragments of up to 23,000 bp

102 Replication of phage- λ DNA in lytic and lysogenic cycles

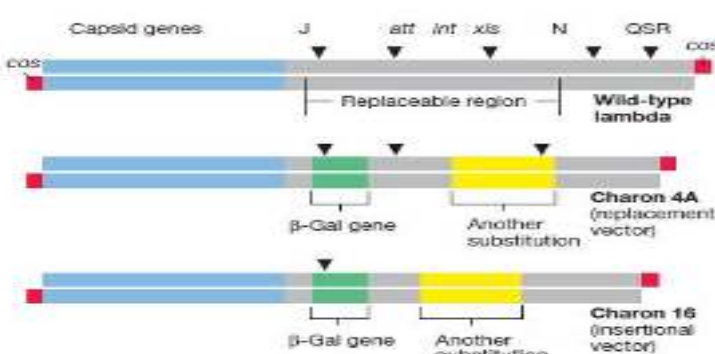
The lytic cycle results in the destruction of the infected bacterial cells

Viral DNA exists as separate molecule within the bacterial cell and replicates independently of the host DNA. Lysogeny is characterized by the integration of the λ DNA into the host bacterium genome.



103 Modified λ phages

Wild type is not suitable as a cloning vector because it has too many restriction enzymes sites



Insertional vectors

Replacement vectors

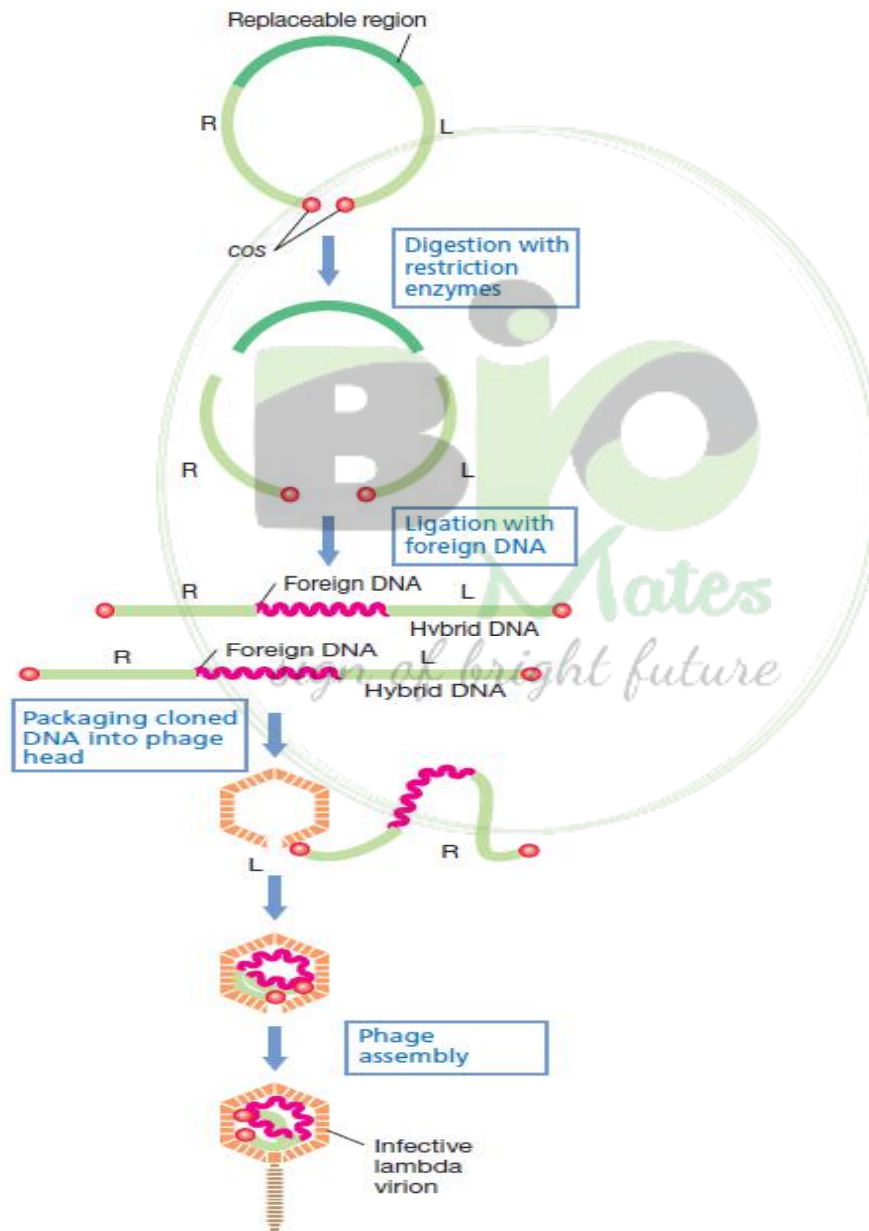
Charon 4A and 16 are both derivatives of lambda with various substitutions and deletions in the nonessential region

Each has the *lacZ* gene, encoding enzyme β -galactosidase, which permits detection of phage containing inserted clones

104 Steps in cloning with λ

Cloning with lambda replacement vectors involves the following steps

- (i). Isolation and cutting vector DNA with appropriate restriction enzymes
- (ii). Connecting the two lambda fragments to foreign DNA by using DNA ligase
- (iii). In vitro packaging of recombinant DNA
- (iv). Infection of *E. coli* cells and isolation of phage clones by picking plaques on a host strain
- (v). Checking recombinant phage for the presence of foreign DNA

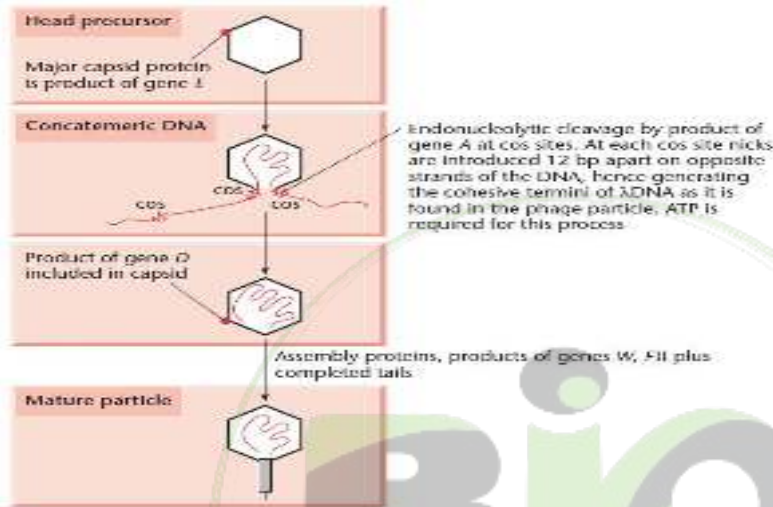


Both Charon vectors are also engineered to contain reporter genes such as for β -galactosidase for detection of recombinant phages

105 Packaging phage- λ DNA in vitro

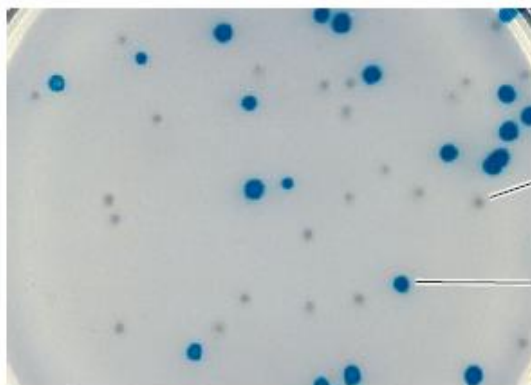
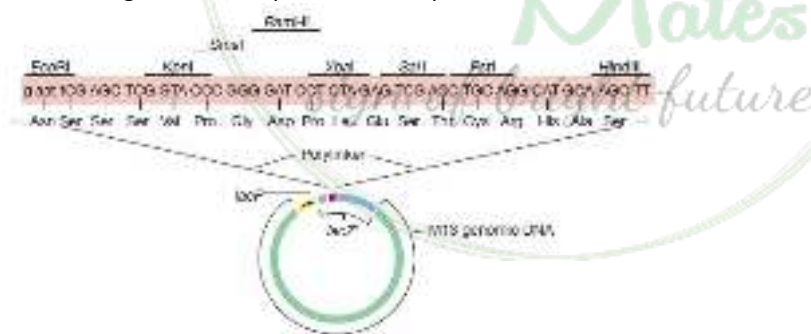
Recombinant DNA in a phage coat allows it to be introduced into the host bacteria by the normal processes of phage infection

Packaging in vitro yields about 10^6 plaques/ μg of vector DNA



106 Vectors for DNA sequencing: bacteriophage M13

M13 is a filamentous bacteriophage with single stranded DNA that replicates without killing its host
Variable lengths of DNA upto about 5 kbp, can be cloned



Phage in clear plaques have cloned DNA

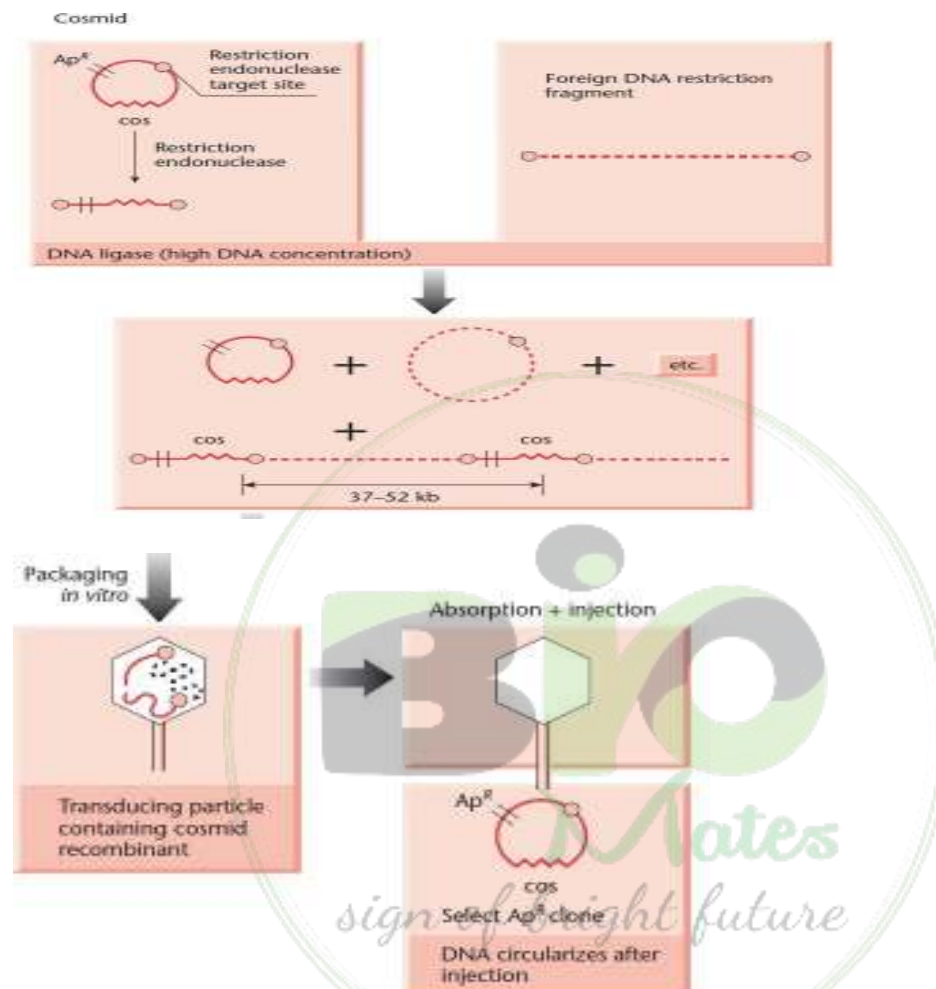
Phage in blue plaques do not have cloned DNA

Use of M13 in molecular cloning

107 Cosmid vectors

Plasmids have been constructed which contain a fragment of λ DNA including 'cos' site

These plasmids have been termed cosmids and can be used as gene-cloning vectors in conjunction with the in vitro packaging system



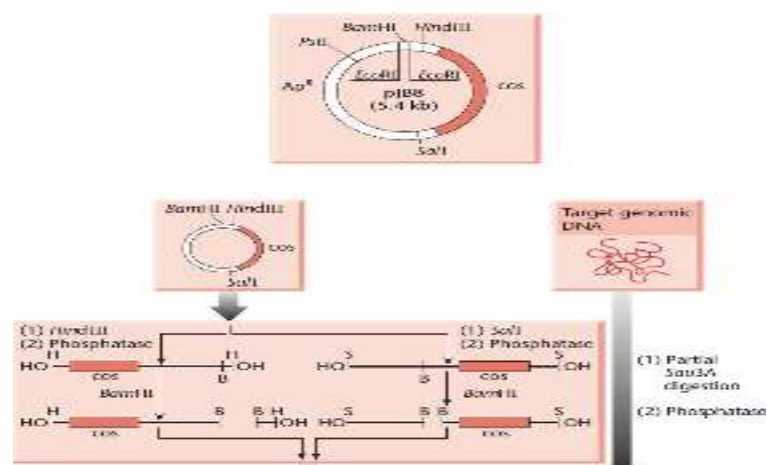
Cosmid vectors

Cosmids provide an efficient means of cloning large pieces of foreign DNA

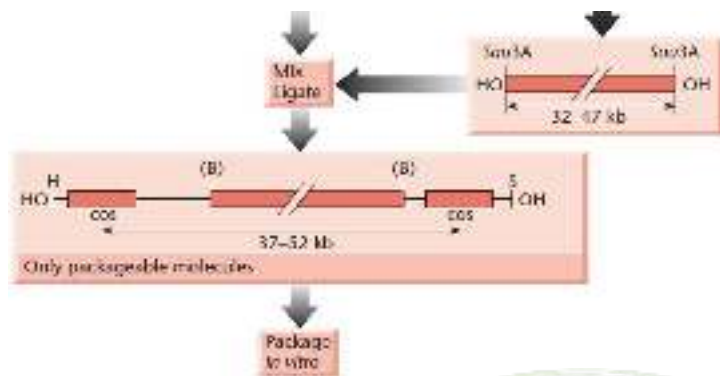
Cosmids are particularly attractive vectors for constructing libraries of eukaryotic genomic fragments

108 Modified schemes for cloning in Cosmid vectors

Difficulties in cloning with cosmids have been overcome in a cosmid cloning procedure devised by Ish-Horowicz and Burke (1981)



Plasmid cloning scheme of Ish-Horowicz and Burke (1981)



Plasmid cloning scheme of Ish-Horowitz and Burke (1981)

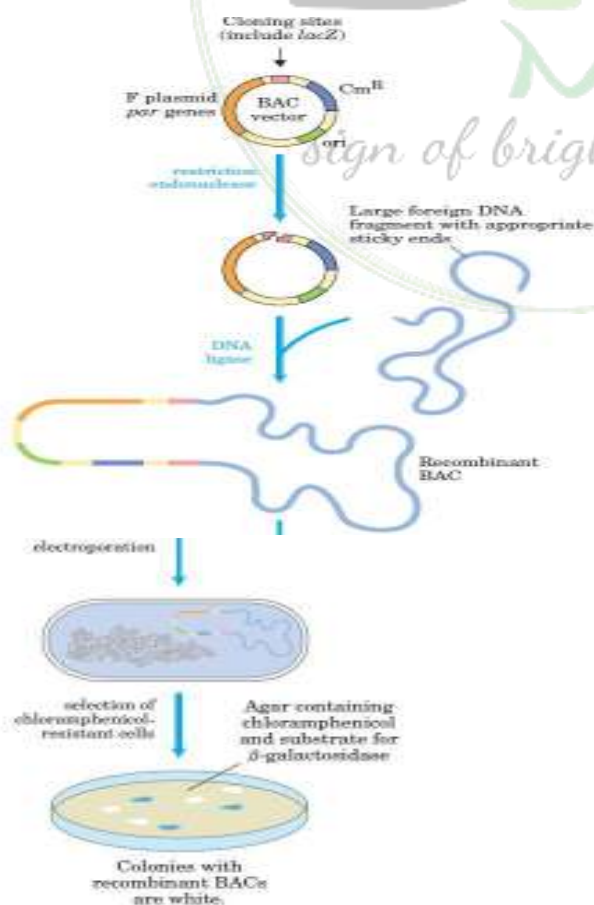
An alternate solution to these problems has been devised by Bates and Swift (1983) Who constructed cosmid c2XB

109 Phasmid vectors

Phasmid vectors are combination of plasmid and λ phage sequences
It consists of a plasmid vector carrying a λ attachment (λ att) site

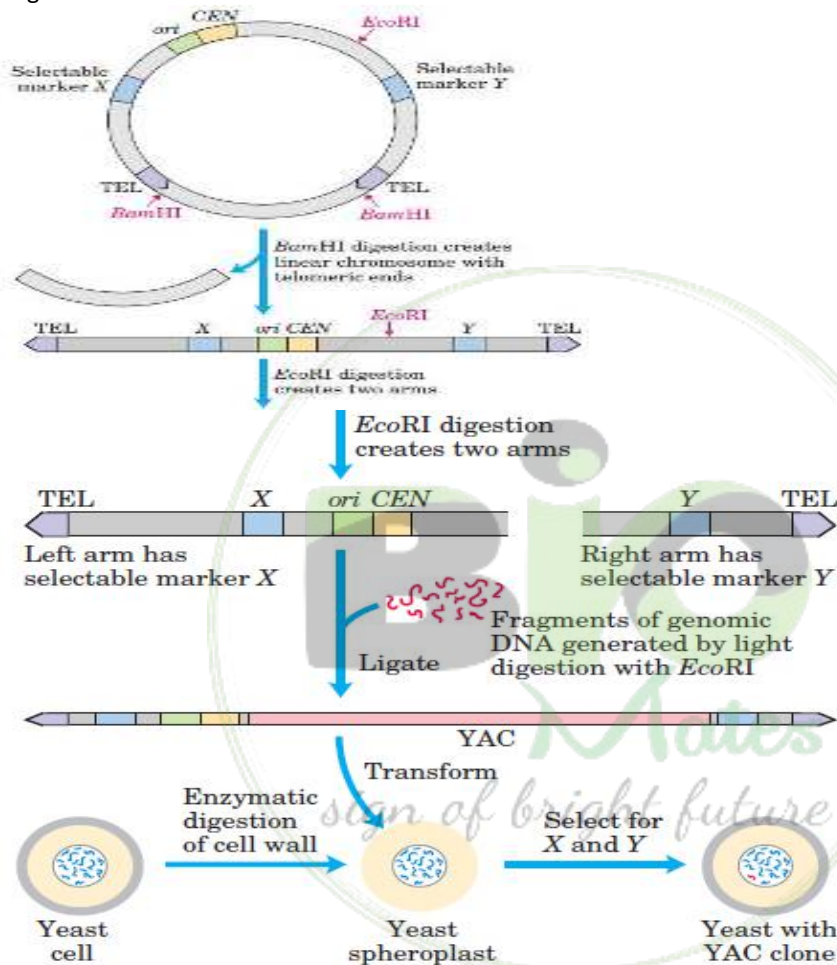
110 Bacterial artificial chromosomes (BACs)

Bacterial artificial chromosomes are simply plasmids designed for the cloning of very long segments of DNA (100,000 to 300,000 bp)



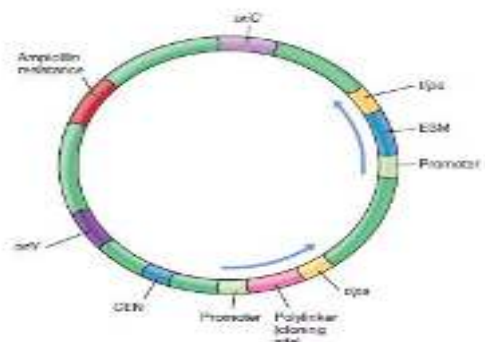
Bacterial artificial chromosome as cloning vectors

These vectors replicate in yeast like normal chromosome but they have sites where very large DNA fragments can be inserted



112 Shuttle and Expression vectors

Vectors that can replicate and are stably maintained in two or more unrelated host organisms are called shuttle vectors



113 Comparison of different cloning vectors-summary

77

Vector	Host	Insert size
pBR322	<i>E. coli</i>	10-15 kb
λ phage	<i>E. coli</i>	2-25 kb
λ Cosmids	<i>E. coli</i>	35-45 kb
P1 phage	<i>E. coli</i>	70-100 kb
PACs	<i>E. coli</i>	100-300 kb
BACs	<i>E. coli</i>	≤ 300 kb
YACs	<i>S. cerevisiae</i>	200-2000 kb

114

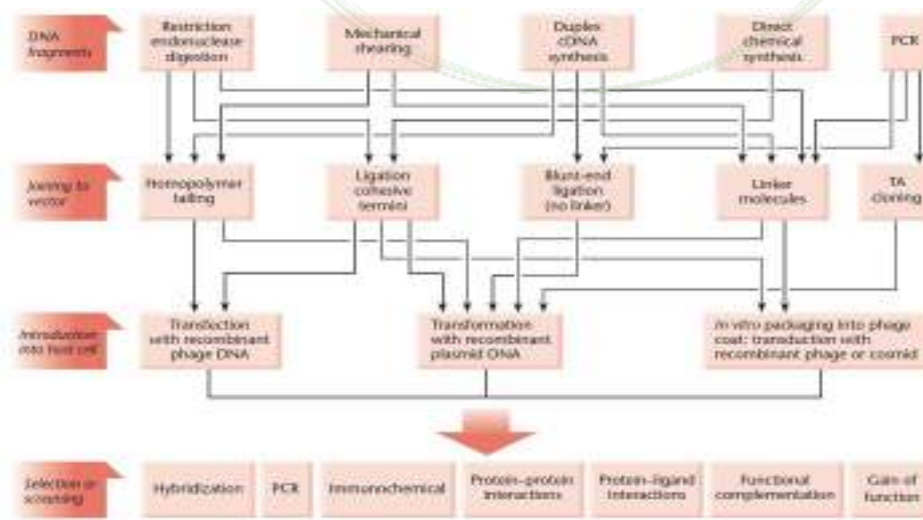
OVERVIEW OF CLONING STRATEGIES

Any cell based cloning procedure has four essential parts

- A method of generating DNA fragment for cloning
- A reaction that inserts that fragment into the chosen cloning vectors

Introduction

- A means for introducing that recombinant vector into a host cell wherein it is replicated
- A method of selecting recipient cells that have acquired the recombinant



Introduction

There are two major strategies for isolating sequences from complex sources

- Cell based cloning strategy, is to divide the source DNA into manageable fragments and clone everything

115 Genomic DNA libraries

Producing representative genomic libraries in λ cloning vector

Organism	Genome size (kb) (haploid where appropriate)
<i>Escherichia coli</i>	4.0×10^3
Yeast (<i>Saccharomyces cerevisiae</i>)	1.35×10^4
<i>Arabidopsis thaliana</i> (higher plant)	1.25×10^5
Tobacco	1.6×10^6
Wheat	5.9×10^6
<i>Zea mays</i>	1.5×10^7
<i>Drosophila melanogaster</i>	1.8×10^5
Mouse	2.3×10^6
Human	2.8×10^6
<i>Xenopus laevis</i>	3.0×10^6

Genomic DNA libraries

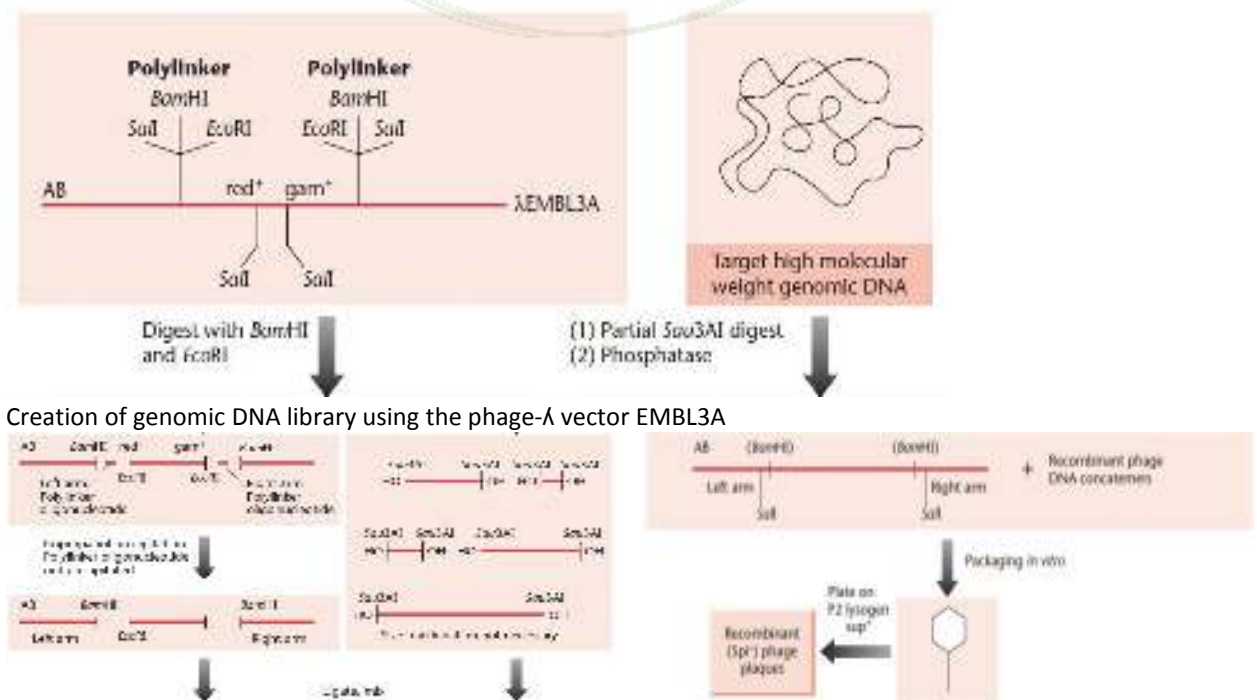
Cloning of large DNA fragments of 20 kb in λ replacement vectors, fewer clones are required for nearly complete library

How can appropriately sized random fragments be produced?

- Average fragment size can be controlled by mechanical shearing
- The more commonly used procedure is restriction endonucleases

116 λ EMBL vectors for genomic library construction

High molecular weight genomic DNA is digested with *Sau3A*I and subsequently ligated in λ EMBL vector digested with *Bam*HI



117 Genomic libraries in high-capacity vectors

In place of phage- λ derivatives, a number of high capacity cloning vectors such as cosmids, bacterial artificial chromosomes (BACs) and yeast artificial chromosomes (YACs) are available for construction of genomic libraries

118 PCR as an alternative to genomic DNA cloning

PCR with specific primers could be used to isolate genes directly from genomic DNA, obviating the need for the production of genomic libraries

119 Properties of cDNA

cDNA is prepared by reverse-transcribing cellular RNA

cDNA lack introns and other non-coding sequences present in the corresponding genomic DNA

Introns are rare in bacteria but occur in genes of higher eukaryotes

120 cDNA libraries

cDNA library is a combination of cloned cDNA fragments inserted into a collection of host cells

cDNA library is representative of the RNA population from which it was derived

Source	Number of different mRNAs	Abundance (molecules/cell)
Mouse liver cytoplasmic poly(A) ⁺	9	12 000
	700	300
	11 500	15
Chick oviduct polysomal poly(A) ⁺	1	100 000
	1	4 000
	12 500	5

References: mouse (Young et al. 1976); chick oviduct (Axel et al. 1975).

cDNA libraries

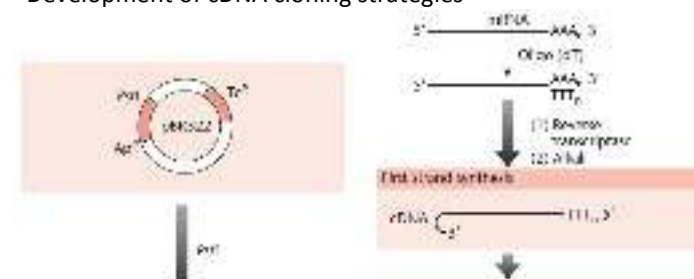
Phage- λ vectors for cDNA cloning and expression

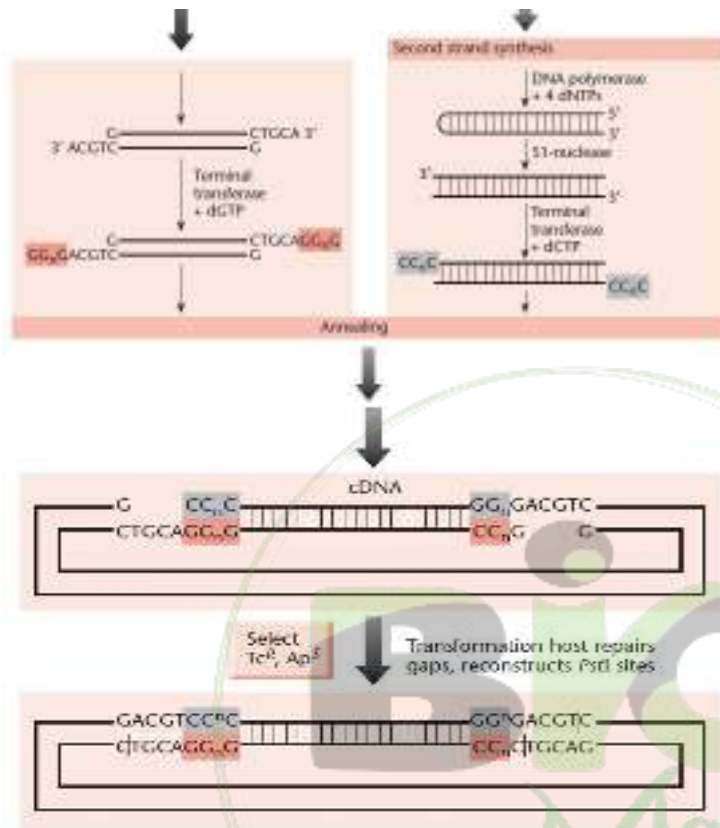
λ gt10 and λ gt11 vectors

121 Preparation of cDNA for library construction

The cDNA synthesis reaction

Development of cDNA cloning strategies

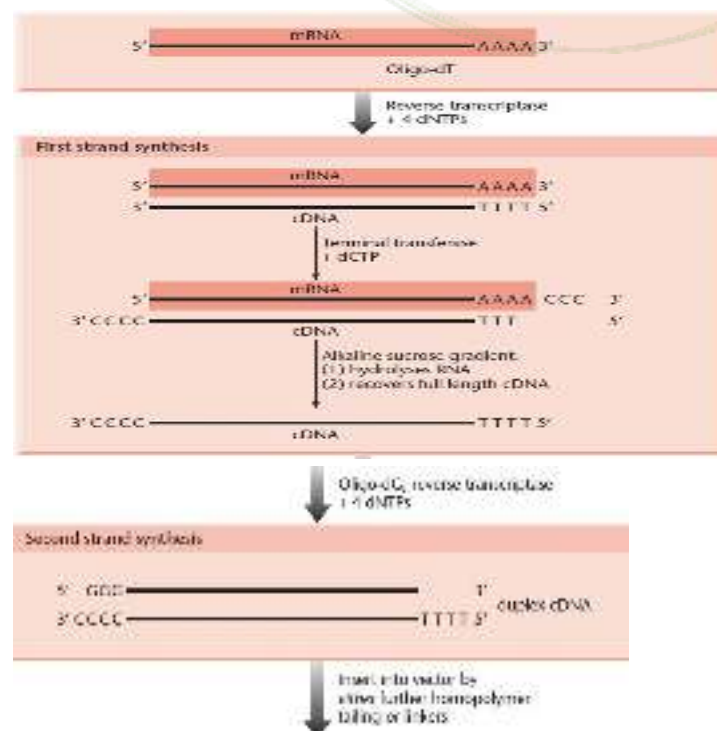




122 Improved methods for cDNA cloning

A serious disadvantage of the hairpin method is that cleavage with S1 nuclease results in the loss of sequences at the 5' end of the clone

This strategy has therefore been superseded with other methods



For cDNA expression libraries, it is advantageous if the cDNA can be inserted into the vector in the correct orientation

This can be achieved by self-priming method by adding linker molecule to double stranded cDNA

123 PCR as an alternative for cDNA cloning

Reverse transcription followed by the PCR (RT-PCR) leads to the amplification of RNA sequences in cDNA form

RT-PCR is a sensitive means for detecting, quantifying and cloning specific cDNA molecules

124

Major screening strategies involve

- 1). Genetic methods
- 2). Sequence-dependent screening
- 3). Screening expression libraries

All useful vector molecules carry a selectable genetic marker or property

Plasmid and cosmid vectors carry drug resistance or nutritional marker

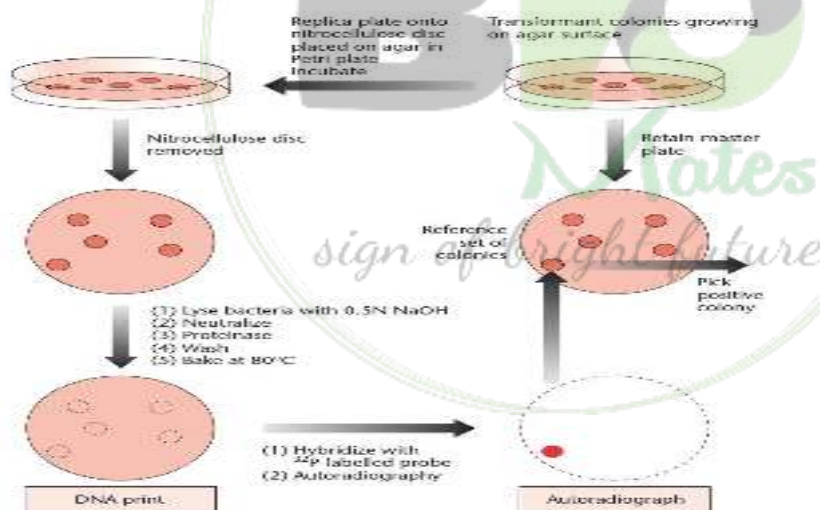
In phage vectors, plaque formation is itself the selected property

125 Sequence-dependent screening

Screening by hybridization

Nucleic acid hybridization is the most commonly used method of library screening

Grunstein and Hogness (1975) developed a screening procedure to detect DNA sequences in transformed colonies by in situ hybridization



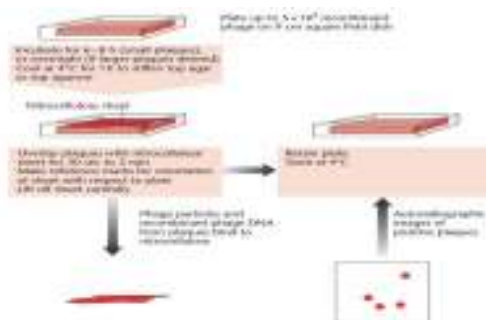
Screening by hybridization

The results of the hybridization can be monitored by autoradiography

126 Sequence-dependent screening

Benton and Davis' plaque lift procedure

Benton and Davis (1977) devised a method called plaque lift, in which nitrocellulose filter is applied to the upper surface of agar plates, making direct contact between plaques and filter



Benton and Davis' plaque lift procedure

Number of alternative labeling methods are available that avoid the use of radioactivity. It included digoxigenin or biotin

127 Probe design

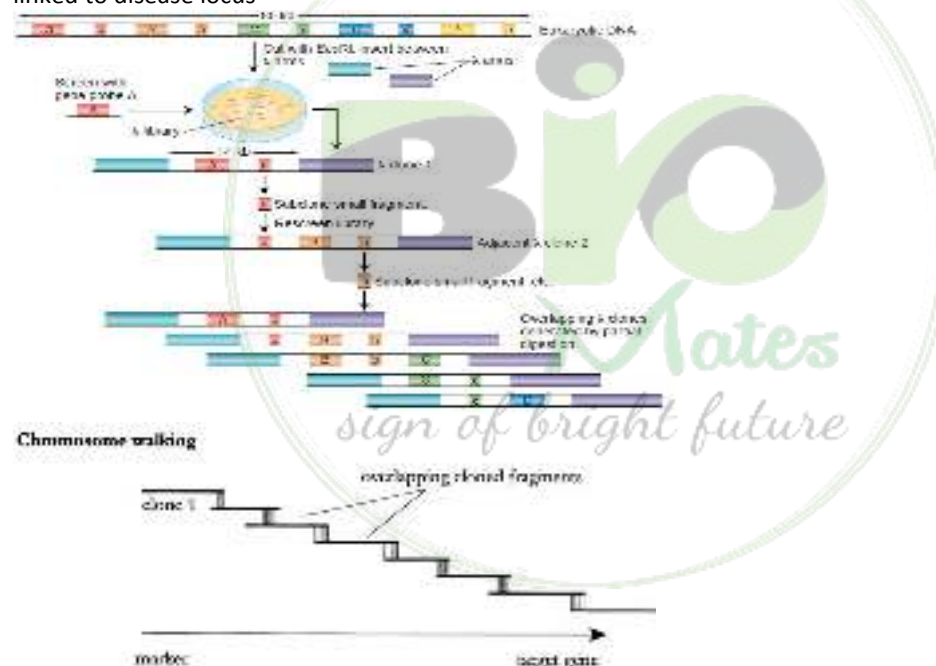
A great advantage of hybridization for library screening is that it is extremely versatile

Conditions can be used in which hybridization is very stringent, so that only sequences identical to the probe are identified

128 Chromosome walking

Walking along the chromosome is a term used to describe an approach which allows the isolation of gene sequences whose function is quite unknown but whose genetic location is known

A cloned genomic fragment must be found as a starting point for the walk In human genome, the starting point may be a restriction fragment length polymorphism (RFLP) sequence that is closely linked to disease locus

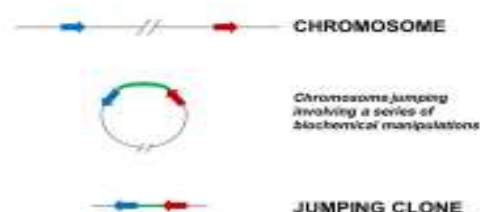


129 Chromosome jumping

In chromosome jumping, the DNA of interest is identified, cut into fragments with restriction enzymes and circularized

It brings together DNA sequence that were originally located a considerable distance apart in the genome

These cloned DNAs from the closure sites make up a jumping library



One of the application of chromosome walking and jumping is in the cloning of human cystic fibrosis gene

130 Screening by PCR

The PCR is widely used to isolate specific DNA sequences from uncloned genomic DNA or cDNA but it also a useful technique for library screening

To isolate specific clone, PCR is carried out with gene-specific primers that flank a unique sequence in the target

There are several applications where the use of degenerate primes is favorable

A degenerate primer is a mixture of primers, all of similar sequences but with variations at one or more positions

131 Expression cloning

If DNA library is established using expression vectors, each individual clone can be expressed to yield a polypeptide

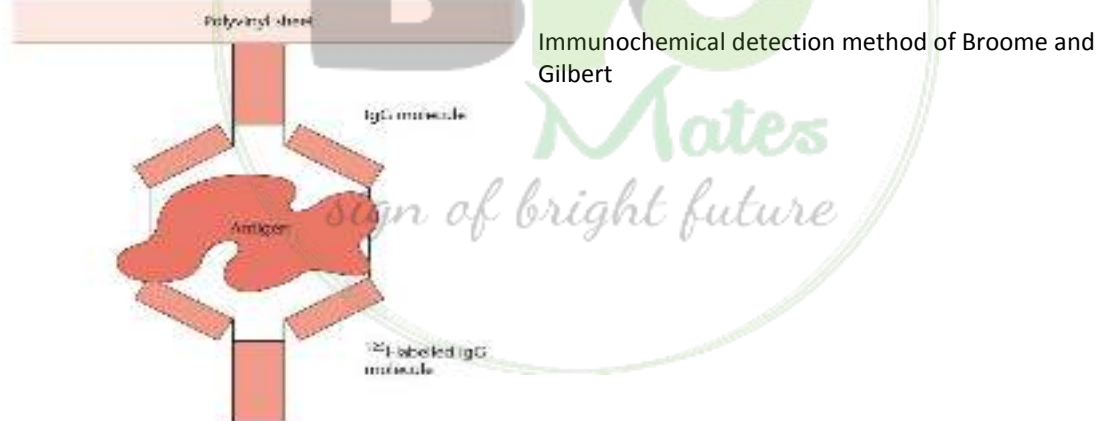
Expression libraries are useful because they allow a range of alternative techniques to be employed

132 Immunochemical screening

Immunochemical screening involves the use of antibodies that specifically recognize antigenic determinants on the polypeptides synthesized by a target clone

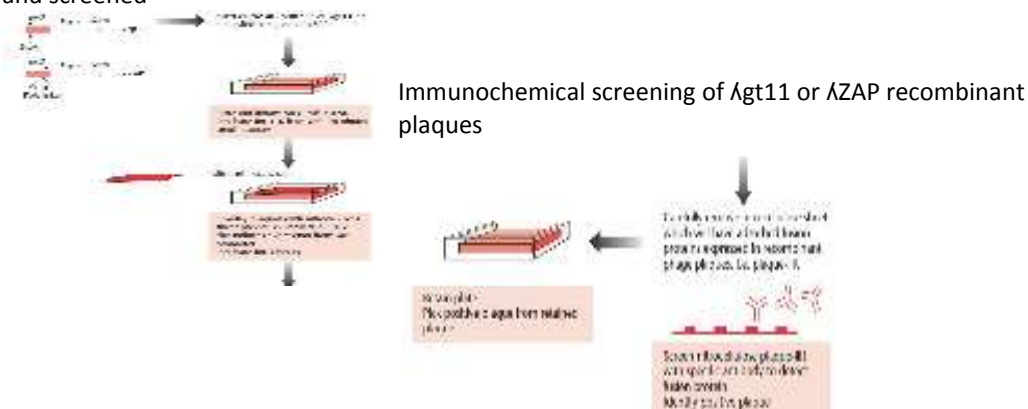
The molecular target for recognition is generally an epitope

The method of Broome and Gilbert (1978) exploited the fact that antibodies adsorb very strongly to certain types of plastic such as polyvinyl and that IgG antibodies can be readily labelled with ¹²⁵I by iodination in vitro



133 Immunochemical screening of λ gt11

It is much more convenient to use bacteriophage- λ insertion vectors, because these have a higher capacity and the efficiency of in vitro packaging allows large numbers of recombinants to be prepared and screened



The original detection method using iodinated antibodies has been superseded by more convenient methods using non-isotopic labels, which are also more sensitive and have a lower background of non-specific signal

134 South-western and north-western blotting

It involves the screening and isolation of clones expressing sequence specific DNA or RNA-binding proteins

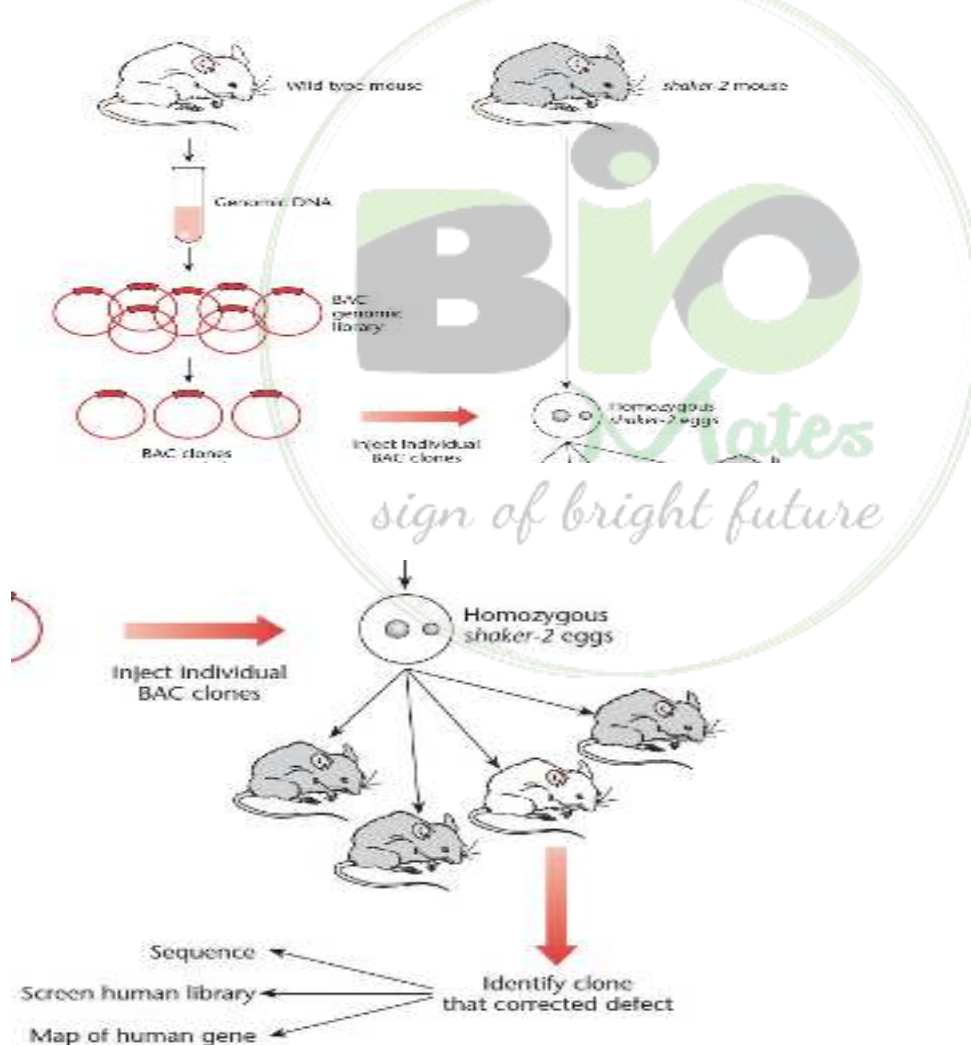
Screening is carried out without using an antibody, by incubating the membranes with radiolabelled double stranded DNA probe

135 Screening by functional complementation

Functional complementation is the process by which a particular DNA sequence compensates for a missing function in a mutant cell and thus restores the wild type phenotype

Functional complementation is also possible in transgenic animals and plants

It has been used for complementation in transgenic mice to isolate the Shaker-2 gene



136 Requirement for expression in E. coli

Synthesis of functional protein depends upon transcription of appropriate gene, efficient translation of mRNA and in many cases, post-translational processing and compartmentalization of nascent polypeptide

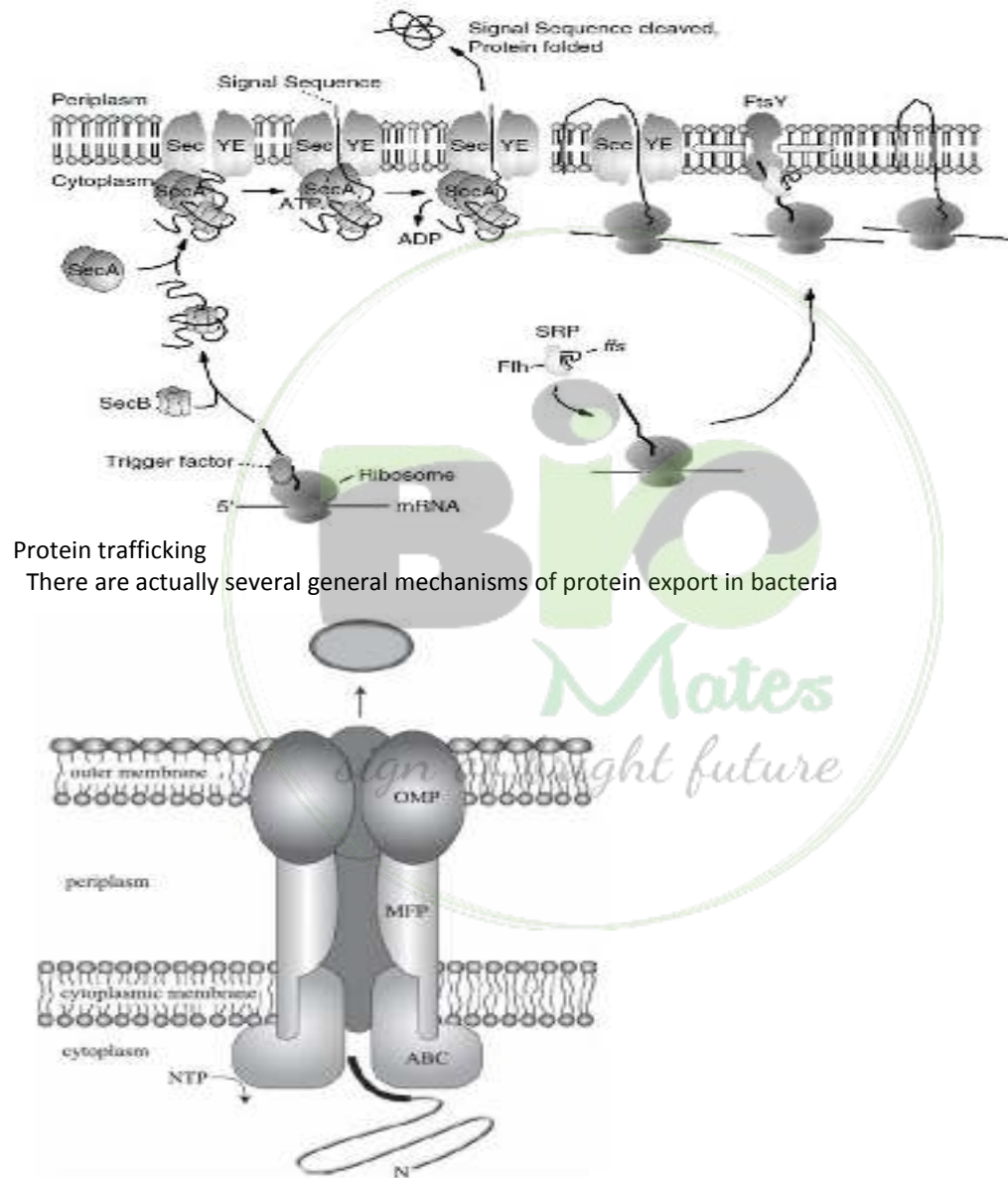
137 Secretion of proteins

Gram-negative bacteria such as *E. coli* have a complex wall-membrane structure comprising an inner, cytoplasmic membrane separated from an outer membrane by a cell wall and periplasmic space

138 Protein trafficking

The bacterial inner membrane, periplasmic space and outer membrane all contain proteins not found in the cytoplasm

All of these proteins are first synthesized in the cytoplasm but somehow find their way out



Protein trafficking

There are actually several general mechanisms of protein export in bacteria

ABC pathway for protein export

139 Stability of foreign proteins in E. coli

Various strategies have been developed to cope with the instability of foreign proteins in E. coli

In the case of somatostatin, degradation was prevented by producing a fused protein consisting of somatostatin and β -galactosidase

140 Constructing the optimal promoter

Large number of promoters for E. coli have been analysed

Many promoters has led to the formulation of a consensus sequence which consists of the -35 region (5'-TTGACA-) and -10 region or Pribnow box (5'-TATAAT-)

	-35 Region		-10 Region
		1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17	
CONSENSUS	***TTGACA*****		TATAAT**
<i>lac</i>	GGCTTTACACTTTATGCTTCGGGCTCGTATATTGT		
<i>trp</i>	CTGTTGACAATTAATCAT	CGAACTAGTTAACTAG	
λ P _L	GTGTTGACATAAATACCA	CTGGCGGTGATACTGA	
<i>recA</i>	CAC TTGATACTGTATGAA	GCATACAGTATAATTG	
<i>tacI</i>	CTGTTGACAATTAATCAT	CGGCTCGTATAAATGT	
<i>tacII</i>	CTGTTGACAATTAATCAT	CGAACTAGTTTAAATGT	

Figure. The base sequence of the -10 and -35 regions for natural and hybrid promoters
Constructing the optimal promoter

Expression from a strong promoter can represent 20-40% of cloned gene product of total cell protein

141 Optimizing translation initiation

Complementarity of Shine-Dalgarno (S-D) sequence with 16S rRNA can affect the rate of translation
Composition of triplet immediately preceding the AUG start codon also affects the efficiency of translation

142 Stability of mRNA and codon choice

The rate of synthesis of a particular protein will depend on the steady-state of mRNA in the cell
Degradation of mRNA usually proceeds by a combination of endonuclease and 3'-exonuclease attack

143 The effect of plasmid copy number

The number of ribosomes in a cell far exceeds any one class of mRNA
One way of increasing the expression of a cloned gene is to increase the number of the corresponding transcript

144 Plasmid stability

Having maximized the expression of a particular gene it is important to consider what effects this will have on the bacterium harbouring the recombinant plasmid

The loss of plasmids due to defective partitioning is called segregative instability

145 Structural instability

Structural instability of plasmids may arise by deletions or rearrangements of DNA

146 Host cell physiology can affect the level of expression

Factors which will be important include the choice of nutrients and environmental parameters

147 DNA sequencing: Benefits and Applications

DNA sequence information is a prerequisite for planning any substantial manipulation of the DNA
Information is useful to various fields i.e. molecular and evolutionary biology, metagenomics, medicine, forensics etc.

DNA sequencing: Benefits and Applications

Techniques for DNA sequencing became available in the late 1970s

148 Maxam-Gilbert method

This method for DNA sequencing makes use of chemical reagents to bring about base-specific cleavage of the DNA

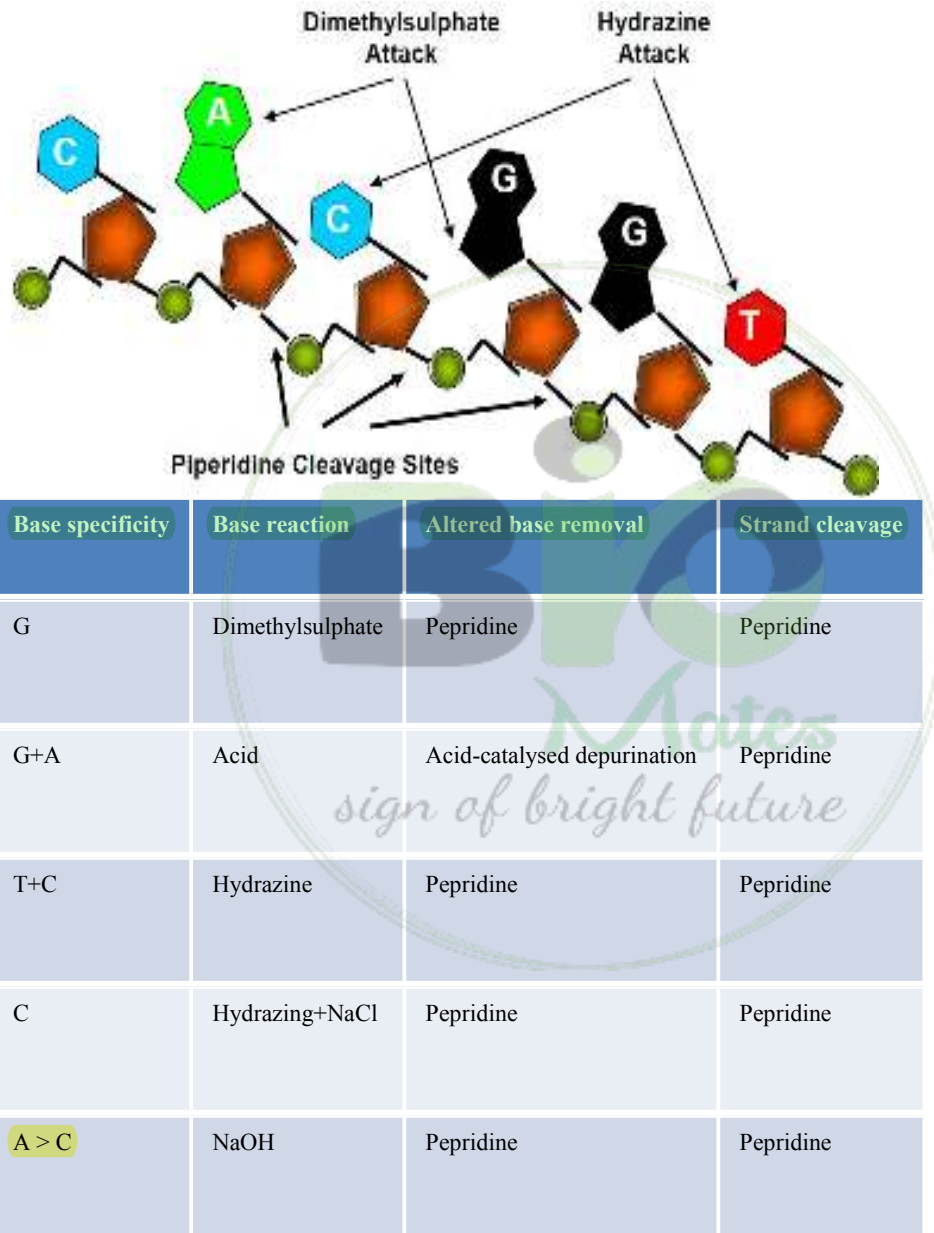
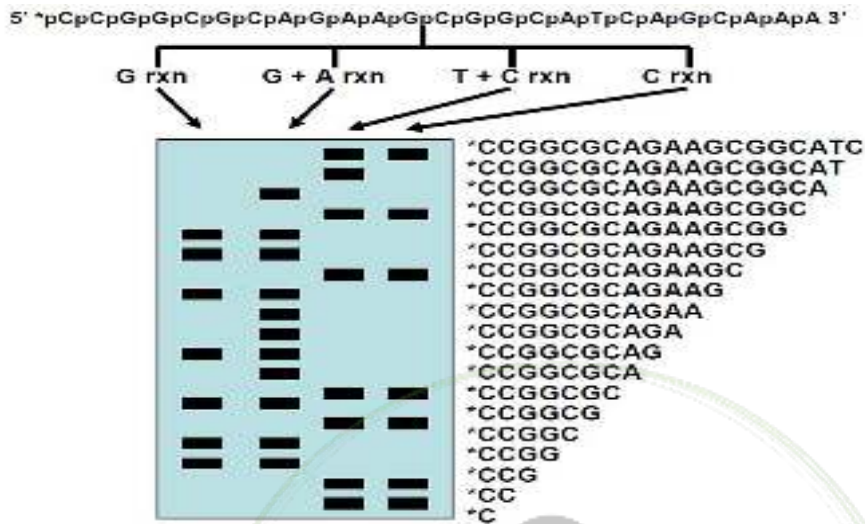


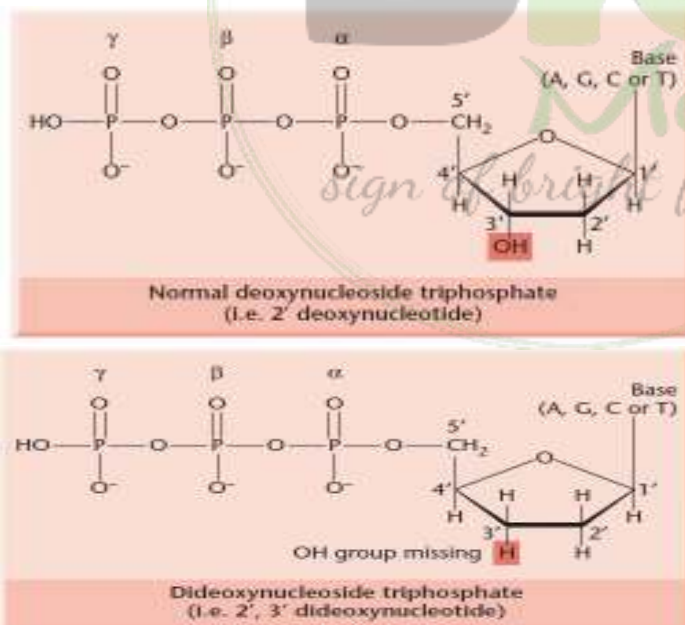
Table. Reagents for Maxam and Gilbert DNA sequencing



149 Chain termination or dideoxy procedure

This method was developed by Sanger et al. (1977)

It capitalizes on the ability of DNA polymerase to incorporate dideoxynucleoside triphosphates as substrate

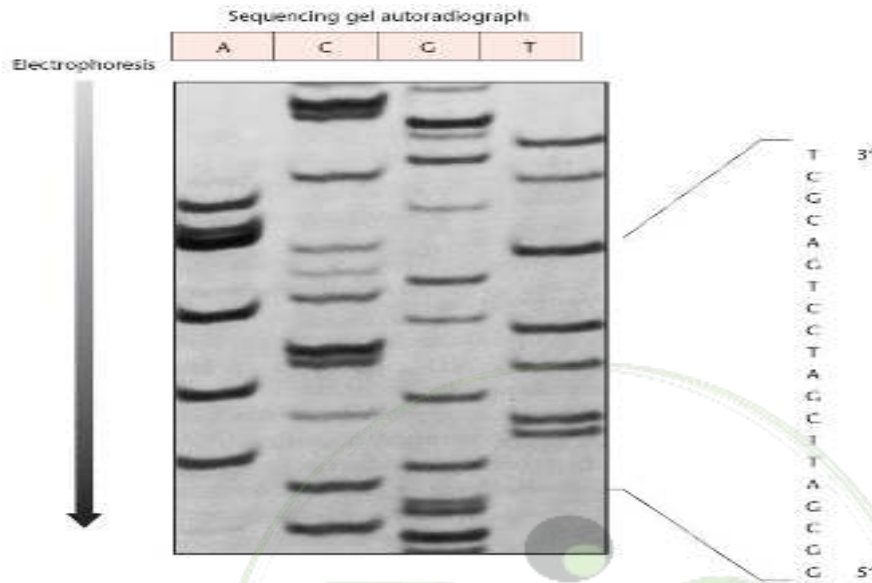


Dideoxynucleoside triphosphate act as chain terminator

Chain termination or dideoxy procedure

DNA synthesis is carried out in the presence of the four deoxynucleoside triphosphates, one or more of which is labelled with ³²P and in four separate incubation mixes





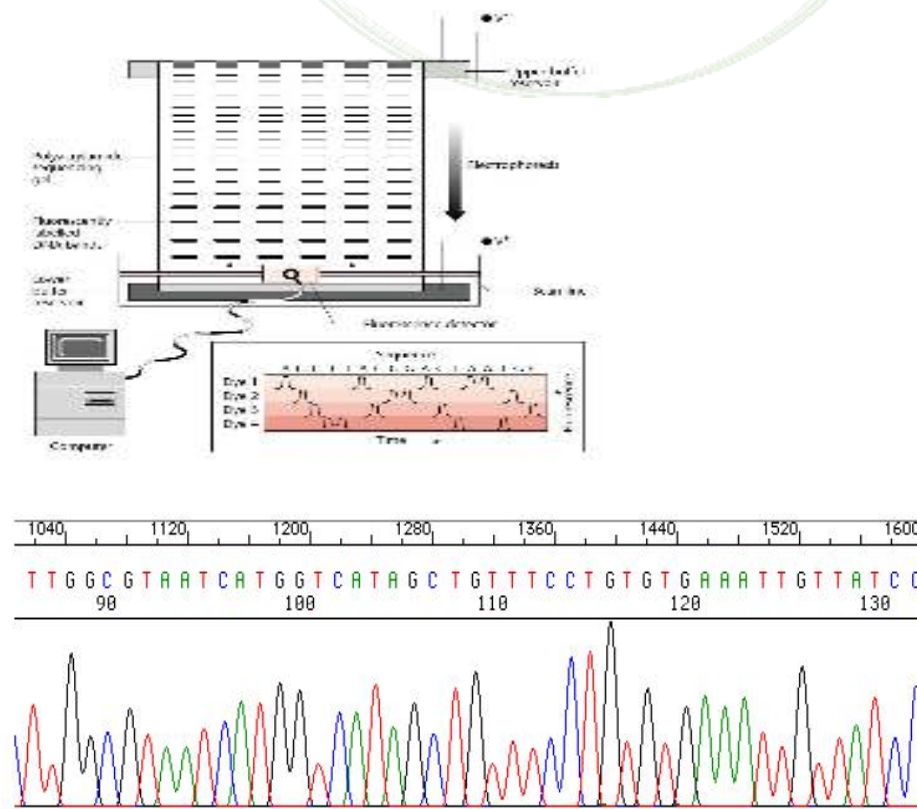
150 Modifications of chain terminator sequencing

Improvements to original Sanger's method have been made by replacing the Klenow fragment of *E. coli* DNA polymerase I

The combination of chain terminator and M13 vectors to produce single stranded DNA is very powerful

151 Automated DNA sequencing

To automate the process, it is desirable to acquire sequence data in real time by detecting the DNA bands within the gel during the electrophoretic separation



Automated DNA sequencing

Automated DNA sequencing offers a number of advantages

152 Sequencing accuracy

Sequences that were read beyond 400 bp contained an average of 3.2% error, while those less than 400 bp had 2.8% error

153 DNA sequence databases

Since the development of current DNA sequencing technology large amount of sequence data has accumulated that is maintained in 3 data bases

- National center for Biotechnology Information
- DNA Databank of Japan
- European Bioinformatic institute-UK

154 Mutagenesis

Mutagenesis is a process to change the genetic information of an organism

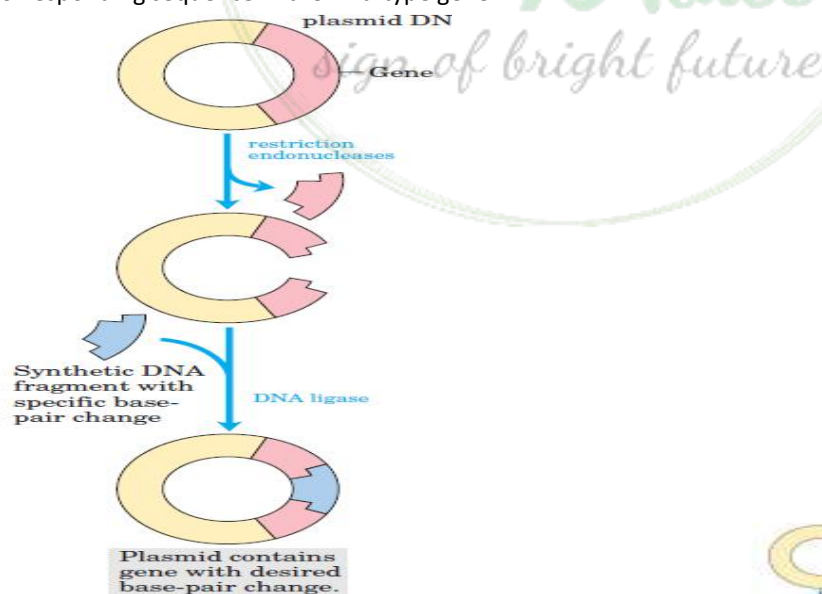
It can occur naturally, or as a result of exposure to mutagens or induced experimentally in laboratory Mutagenesis

Three different methods of site-directed mutagenesis have been devised

- Cassette mutagenesis
- Primer extension
- Procedures based on PCR

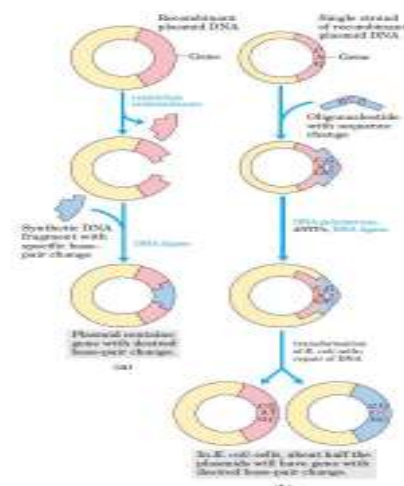
155 Cassette mutagenesis

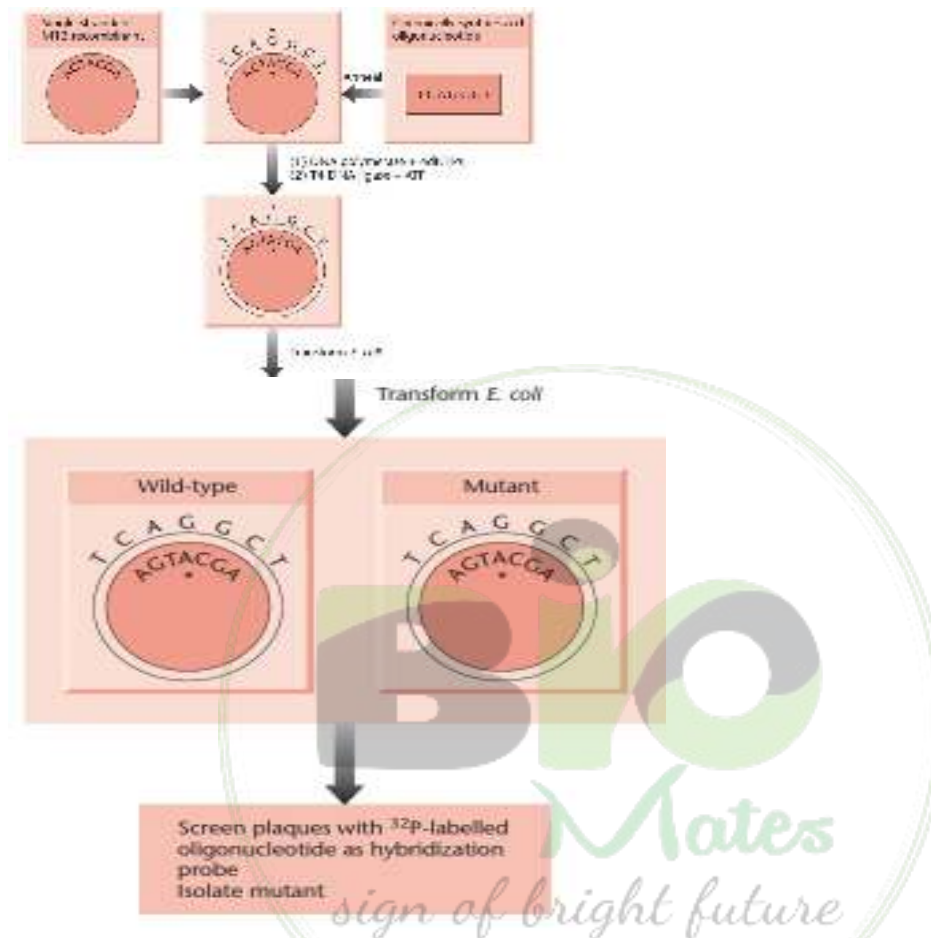
A synthetic DNA fragment containing the desired mutant sequence is used to replace the corresponding sequence in the wild-type gene



156 Primer extension: the single primer method

It involves priming in vitro DNA synthesis with a chemically synthesized oligonucleotide (7-20 bp) that carries a base mismatch with the complementary sequence

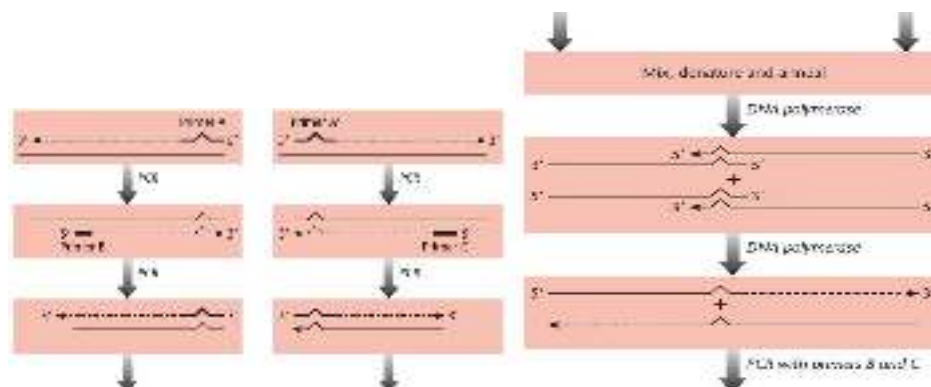




After E. coli transformation, heteroduplex that are formed are either that of original wild type DNA or that containing the mutated base

157 PCR methods for site-directed mutagenesis

Single bases mismatched between the amplification primer and the template become incorporated into the template sequence as a result of amplification



Site directed mutagenesis by means of the PCR

Site directed mutagenesis by means of the PCR

In megaprimer method, the product of the first PCR is used for the second PCR

160 DNA polymerases

DNA polymerases are enzymes that synthesize DNA molecules by adding nucleotide, the building blocks of DNA

Table. Comparison of DNA polymerases of *E. coli*

	DNA polymerase		
	I	II	III
Structural gene*	<i>polA</i>	<i>polB</i>	<i>polC (dnaE)</i>
Subunits (number of different types)	1	7	≥10
M_r	103,000	88,000 [†]	791,500
3'→5' Exonuclease (proofreading)	Yes	Yes	Yes
5'→3' Exonuclease	Yes	No	No
Polymerization rate (nucleotides/s)	16-20	40	250-1,000
Processivity (nucleotides added before polymerase dissociates)	3-200	1,500	≥500,000

Taq DNA polymerase

Taq polymerase is a thermostable DNA polymerase isolated from thermophilic bacterium *Thermus aquaticus* and is used in PCR to amplify DNA fragments

DNA polymerase	Source
<i>Tma</i>	<i>Thermotoga maritima</i>
Deep Vent™	<i>Pyrococcus</i> sp.
<i>Tli</i>	<i>Thermococcus litoralis</i>
<i>Pfu</i>	<i>Pyrococcus furiosus</i>
<i>Pwo</i>	<i>Pyrococcus woesei</i>

Table. Sources of thermostable DNA polymerases

162 Degenerate primers

A degenerate primer is actually a mixture of primers, all of similar sequence but with variations at one or more positions

One Codon	Two Codons	Three Codons	Four Codons	Six Codons
Met (M)	Cys (C)	Ile (I)	Ala (A)	Leu (L)
Trp (W)	Asp (D)		Gly (G)	Arg (R)
	Glu (E)		Pro (P)	Ser (S)
	Phe (F)		Thr (T)	
	His (H)		Val (V)	
	Lys (K)			
	Asn (N)			
	Gln (Q)			
	Tyr (Y)			

Symbol	Base represented
A	A
C	C
G	G
T	T
R	A, G
Y	C, T
M	A, C
K	G, T
H	A, C, T
B	C, G, T
V	A, C, G
N	A, C, G, T

“

Degenerate primers

A degenerate primer is actually a mixture of primers, all of similar sequence but with variations at one or more positions

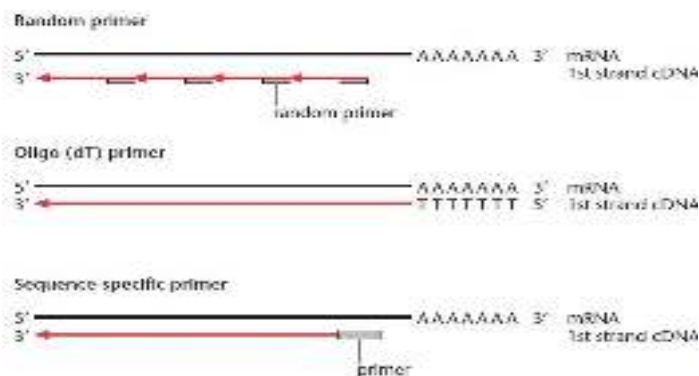
163 Types of PCR

Recently, variants have been developed from basic PCR method to improve performance and specificity and to amplify other molecules such as RNA

RT-PCR

Reverse transcriptase PCR (RT-PCR) was designed to amplify RNA sequences (mainly mRNA) through synthesis of cDNA by reverse transcriptase

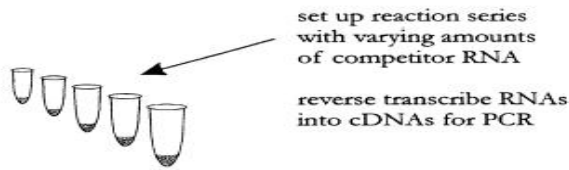
Various strategies can be adopted for first-strand cDNA synthesis



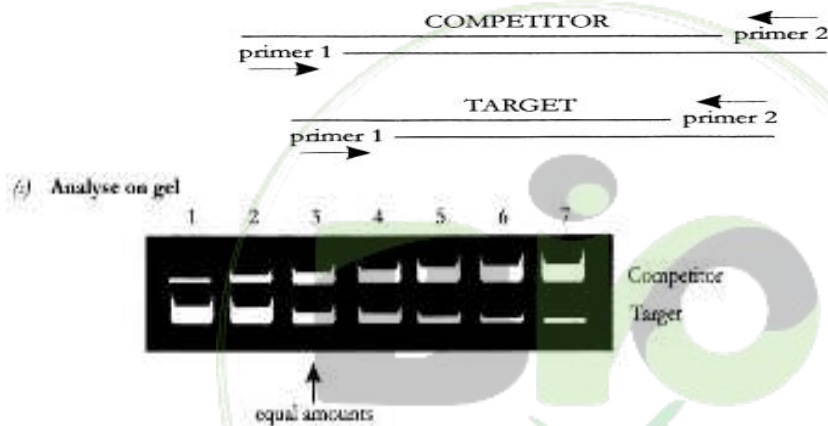
164 Competitor RT-PCR

One use of RT-PCR is in determining the amount of mRNA in a sample (competitor RT-PCR)

(a) Spike RNA samples and convert to cDNA



(b) Perform PCR using same primer pair

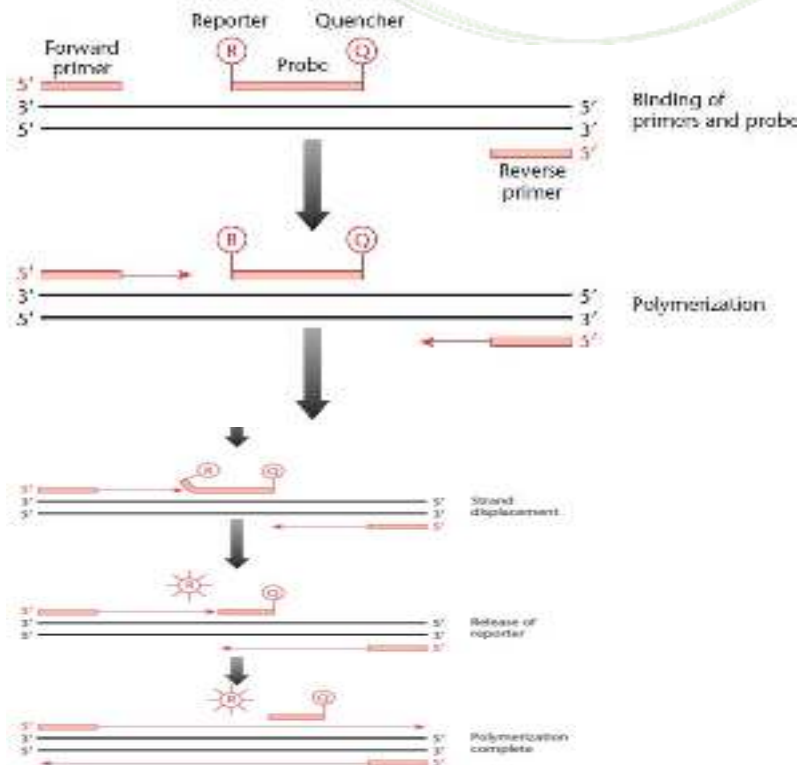


RT-PCR

Reverse transcriptase PCR (RT-PCR) was designed to amplify RNA sequences (mainly mRNA) through synthesis of cDNA by reverse transcriptase

165 Real-time quantitative PCR (qPCR)

qPCR is used to measure the quantity of a target sequence
It is used to measure starting amounts of DNA, cDNA, or RNA



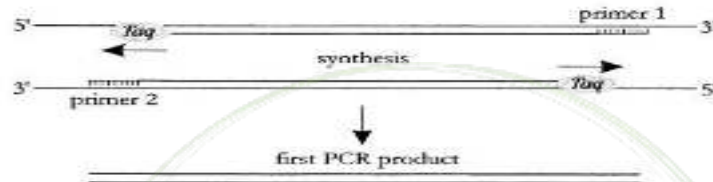
Real-time quantitative PCR (qPCR)

qPCR is used to measure the quantity of a target sequence
It is used to measure starting amounts of DNA, cDNA, or RNA

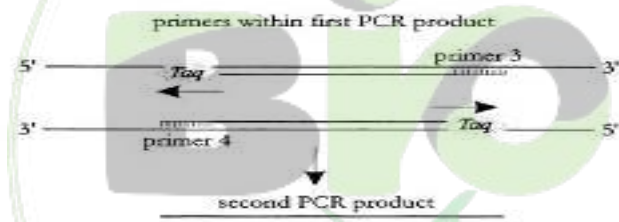
166 Nested PCR

In nested PCR, two sets of primers are used in two successive PCRs
First set of primers allows a first amplification
Product of first PCR is subjected to a second PCR using the second set of primers

(a) First PCR using external primers



(b) Second PCR using internal (nested) primers

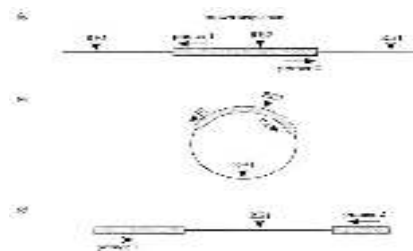


Nested PCR

In nested PCR, two sets of primers are used in two successive PCRs
First set of primers allows a first amplification
Product of first PCR is subjected to a second PCR using the second set of primers

167 Inverse PCR

Inverse PCR is used to identify the flanking sequences around genomic inserts



168 Multiplex PCR

Multiplex PCR is used to amplify several different DNA sequences simultaneously

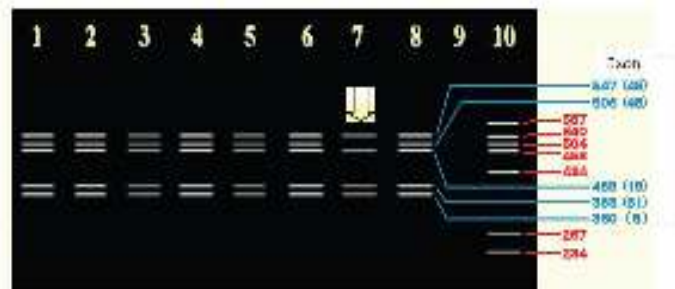
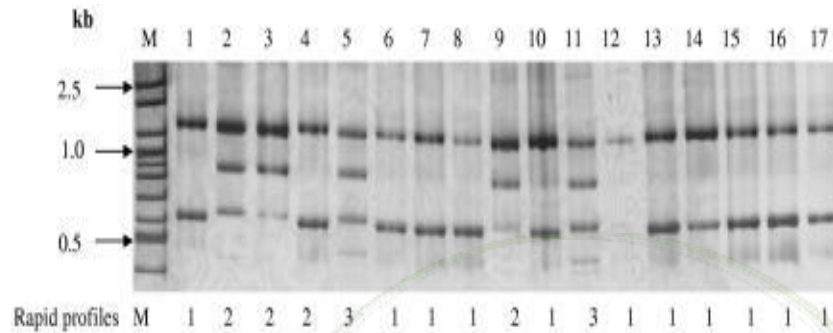


Fig. Results of multiplex PCR in a patient with Duchenne Muscular Dystrophy Inverse PCR
Hernandez-Rodriguez et al., 2000;
Hernandez-Rodriguez & Restrepo, 2002
(www.intechopen.com)

169 RAPD

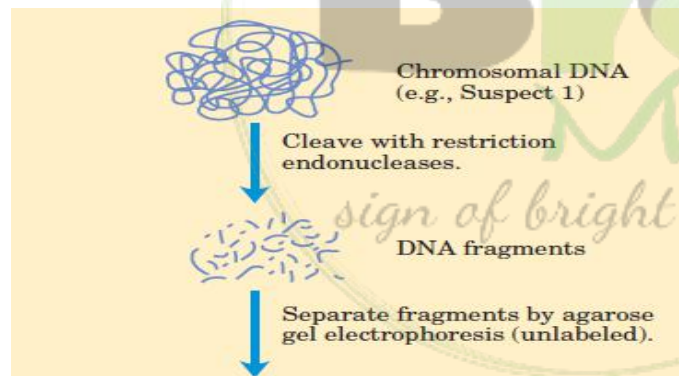
RAPD-Random amplification of polymorphic DNA is a type of PCR reaction but the segments of DNA that are amplified are random



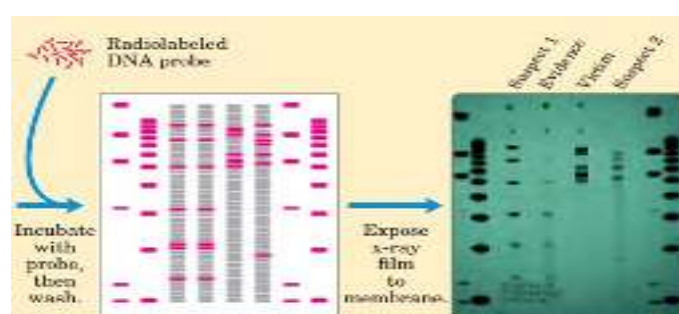
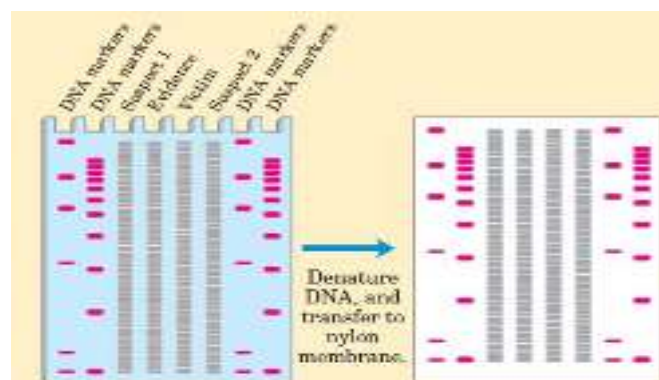
RAPD-Random amplification of polymorphic DNA is a type of PCR reaction but the segments of DNA that are amplified are random

170 RFLP

RFLP-Restriction fragment length polymorphism exploits variations in homologous DNA sequences. In RFLP, DNA sample is digested by restriction enzymes and separated by gel electrophoresis.



DNA fingerprinting



171 AFLP-PCR

AFLPs-Amplified fragment length polymorphisms are differences in restriction fragments length caused by SNPs or INDELS

AFLP is a PCR based tool used in genetic engineering.

172 Applications of PCR

PCR has widespread applications in various fields of life sciences including genetic engineering, medical, forensic, agriculture, environment etc.

PCR has been used in gene cloning and screening of genomic libraries

173 PCR-Medicine

PCR has major impact on medicine especially in the field of clinical microbiology or diagnosis

Molecular tools have also allowed to perform prenatal genetic diagnosis

174 PCR-Forensic sciences

Forensic science is the application of scientific procedures to solve criminal and legal matters

Molecular methods are used to established the filiations of a person or to obtain evidence from minimal samples of saliva, semen or other tissues

PCR-DNA profiling

DNA profiling or DNA fingerprinting is a forensic technique used to identify individuals by characteristics of their DNA

175 PCR-Agricultural sciences and environment

PCR has also facilitated research in detection of pathogens in plants, animals and environment

176 PCR-Molecular paleontology

Molecular paleontology refers to the recovery and analysis of DNA and protein from ancient human, animal and plant remains

177 Genome mapping

Assigning of a specific gene to particular region of a chromosome and determining the location of and relative distances between genes on the chromosome

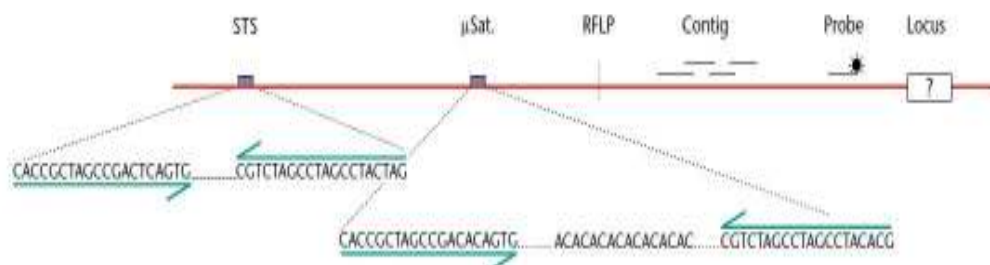
Types of maps:

1. Linkage maps
2. Physical maps

178 Markers for genome mapping

Any identifiable feature of the genome can serve as a marker in mapping

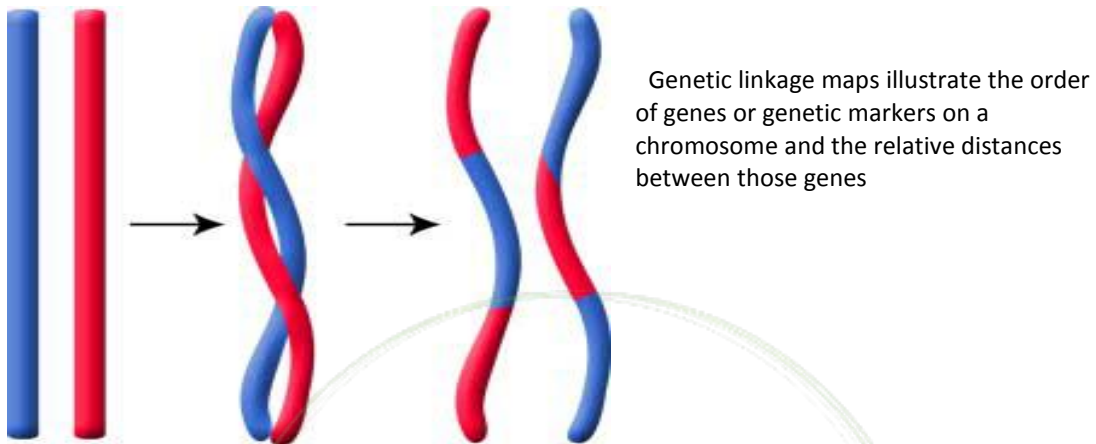
The landmarks on genome map might include short DNA sequences, regulatory sites and genes themselves



179 Genetic linkage mapping

Genetic linkage maps illustrate the order of genes or genetic markers on a chromosome and the relative distances between those genes

Figure. Crossing over in chromosomes
Source: Genomic News Network (GNN)



180 Physical mapping

Physical maps give the DNA base pair distances from one landmark to another

181 Physical versus linkage maps

Chromosomal maps of bacteria has been obtained by combining the mapping techniques of interrupted mating, recombination mapping, transformation and transduction

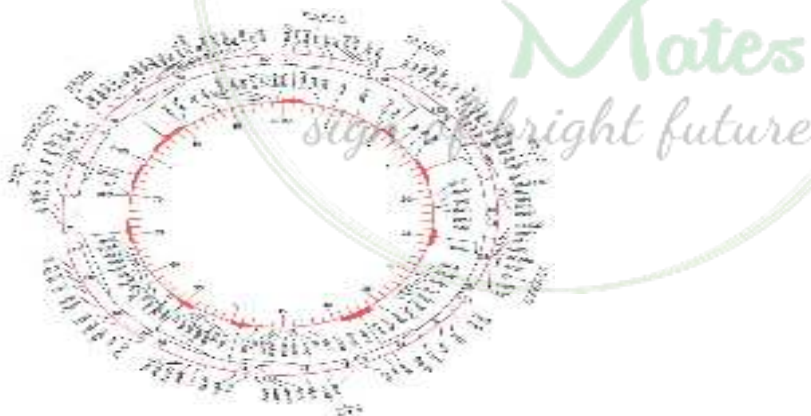


Figure. The 1963 map of E. coli

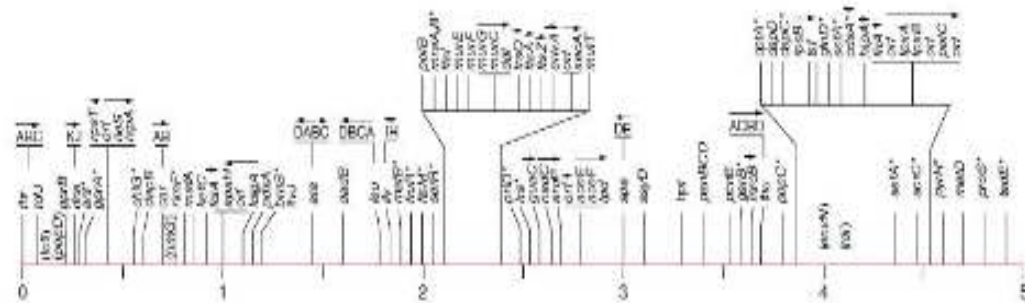


Figure. Linear scale drawing of a 5-minute section of the 100-minute 1990 E. coli linkage map

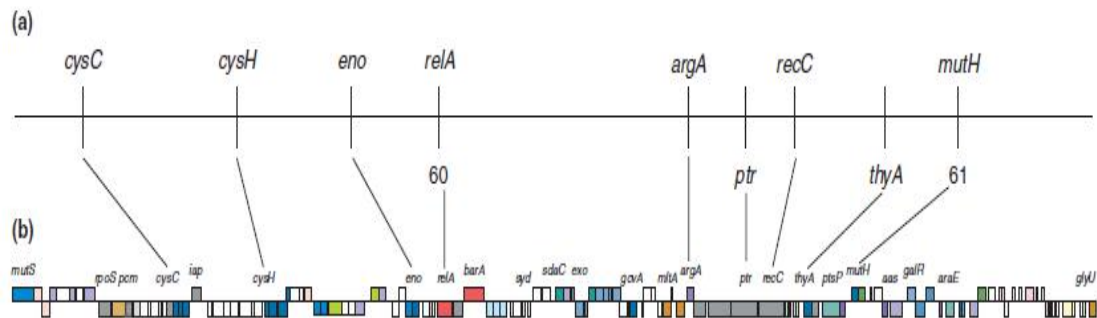


Figure. Correlation of the genetic and physical maps

183 STS in physical maps

Sequence-tagged sites (STS) are more convenient markers than RFLPs because they do not use southern blotting

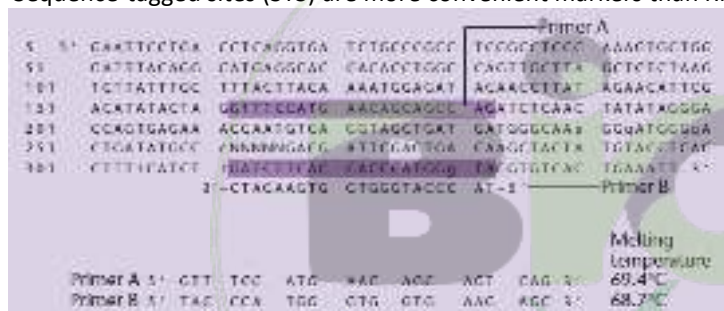
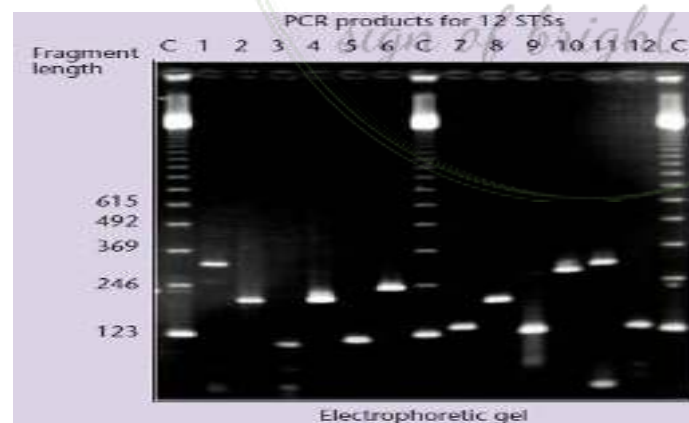


Figure. Examples of a sequence-tagged site (STS)



STS in physical maps

Sequence-tagged sites (STS) are more convenient markers than RFLPs because they do not use southern blotting

184 SNPs as physical markers

Single nucleotide polymorphisms (SNPs) are single base-pair positions in genomic DNA at which different sequence alternatives (alleles) exist in a population

SNPs probably are the most important sequence markers for physical mapping of genomes

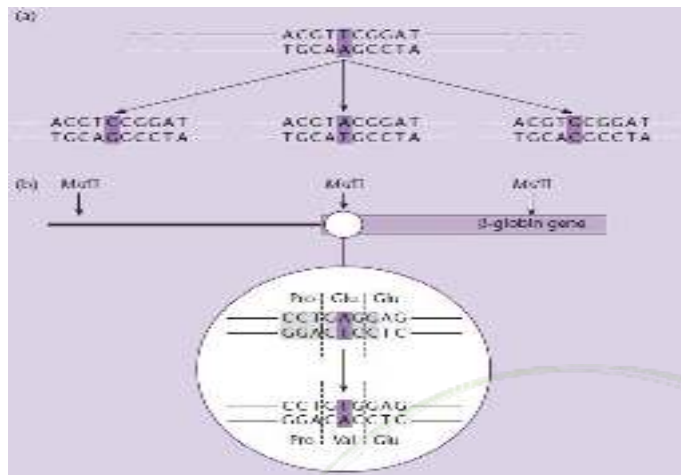


Figure. Examples of single nucleotide polymorphisms SNPs as physical markers

Different methods to detect SNPs have been developed that based on enzymatic, electrophoretic, solid phase or chromatographic analysis

185 Polymorphic DNA detection in the absence of sequence information

Polymorphic DNA can be detected by amplification in the absence of the target DNA sequence information used to generate STSs

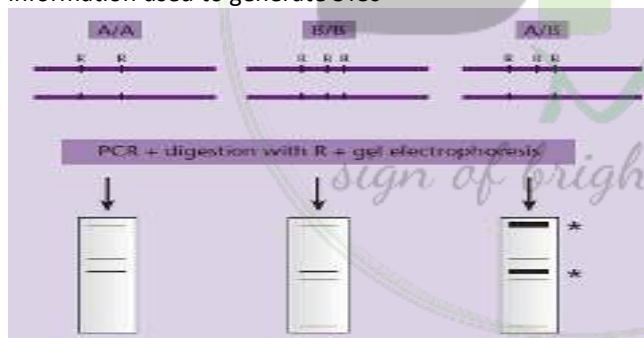
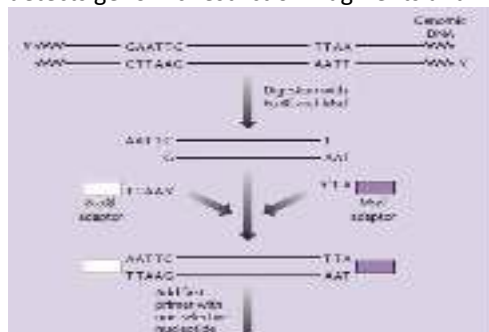


Figure. Generation and visualization of CAPS markers

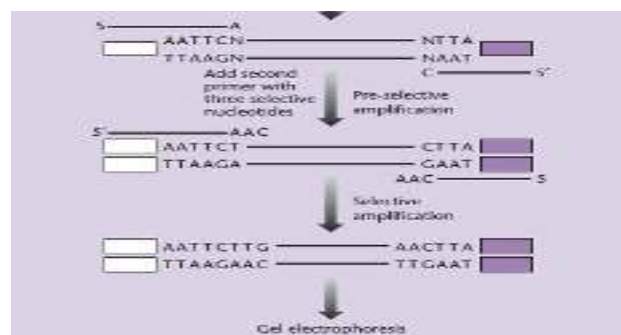
186 AFLPs detection in the absence of sequence information

Amplified fragment length polymorphisms (AFLP) is a diagnostic fingerprinting technique that detects genomic restriction fragments and in that respect resembles the RFLP technique



AFLPs detection in the absence of sequence information

Amplified fragment length polymorphisms (AFLP) is a diagnostic fingerprinting technique that detects genomic restriction fragments and in that respect resembles the RFLP technique



187 Fluorescence in situ hybridization (FISH)

FISH is a cytogenetic method that used fluorescent probes that bind to only those parts of the chromosomes with a high degree of sequence complementarity

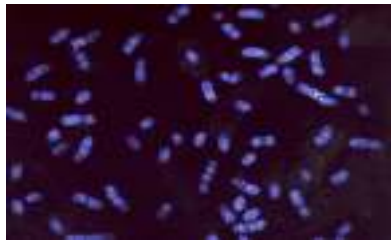


Figure. Dual-color fluorescence in situ hybridization

188 Radiation hybrid (RH) mapping

RH mapping used X-ray breakage of chromosomes to determine the distances between DNA markers as well as their order on the chromosome

189 Happy mapping

RH mapping have a few problems that arise from the in vivo step. Happy mapping overcomes these limitations, being an entirely in vitro technique

190 Map integration

It is essential that the different mapping methods are integrated as each has its own strengths and weaknesses

191 Sequencing genome

The process of determining the exact order of nucleotides within a DNA or RNA molecule is called sequencing

In 1970s, DNA sequencing was a very tedious process which involved determining only a few hundred nucleotides at a time

192 High-throughput sequencing

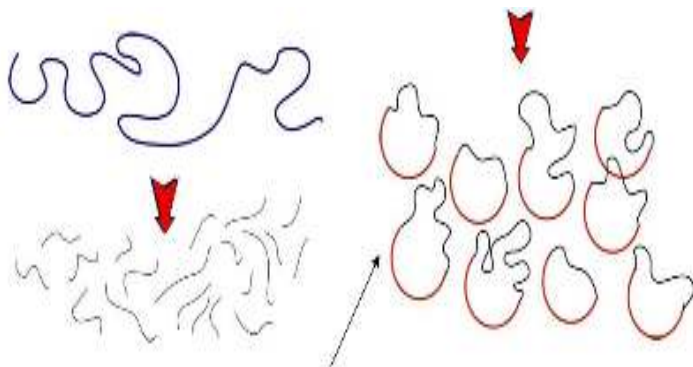
Recent advances in DNA sequencing have make it possible to sequence data very rapidly and at a substantially lower cost

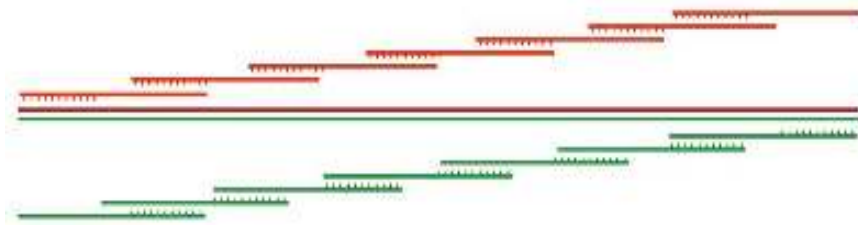
It mainly describe a number of different modern sequencing technologies to sequence genome

193 High-throughput sequencing

Recent advances in DNA sequencing have make it possible to sequence data very rapidly and at a substantially lower cost

It mainly describe a number of different modern sequencing technologies to sequence genome





Shotgun library

194 Clone-by-clone sequencing

In clone-by-clone sequencing, a map of each chromosome of the genome is made before the DNA is split up into fragments for sequencing

196 Comparative genomics of bacteria

Genome of bacteria may be of variable in size. For example, it may vary from 0.49 Mb (Nanoarchaeum equitans) to 9.1 Mb (Bradyrhizobium japonicum or Streptomyces)

197 Comparative genomics of organelles

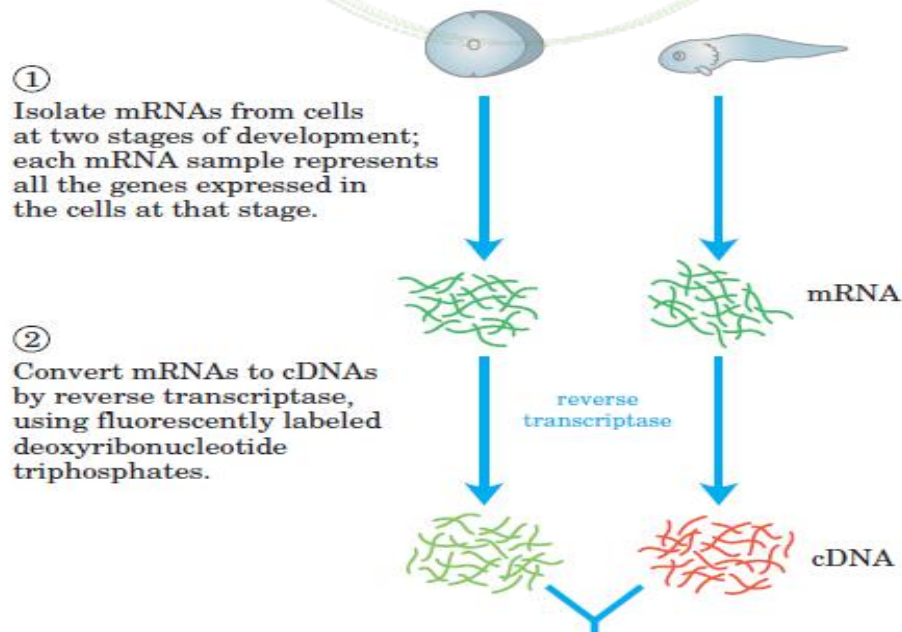
Mitochondrial genomes exhibit an amazing structural diversity
Gene transfer has occurred between mtDNA and nuclear DNA

198 Comparative genomics of eukaryotes

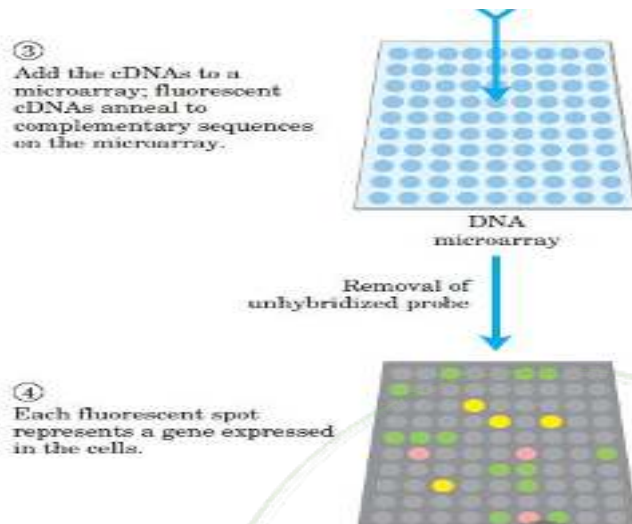
The minimal eukaryotic genome is smaller than many bacterial genomes
Comparative genomics can be used to identify genes and regulatory elements
It gives insight into the evolution of key proteins

199 Comparative genomics of eukaryotes

The minimal eukaryotic genome is smaller than many bacterial genomes
Comparative genomics can be used to identify genes and regulatory elements
It gives insight into the evolution of key proteins



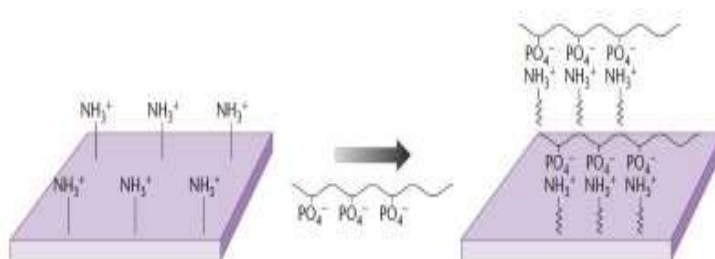
DNA Microarray



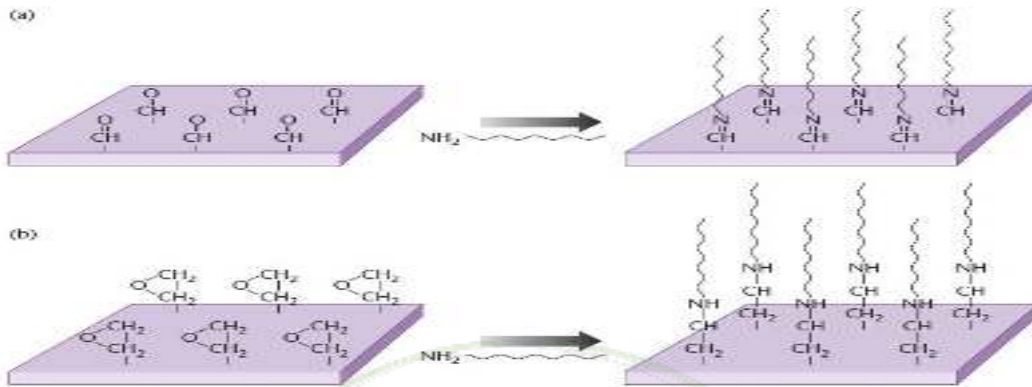
Enlarged image of a DNA Microarray

200 Spotted DNA arrays

Spotted DNA arrays are produced by printing DNA samples on treated microscope slides. It is made by transferring or spotting DNA clones or PCR product individually onto a solid support where they are immobilized



DNA binding by electrostatic interactions



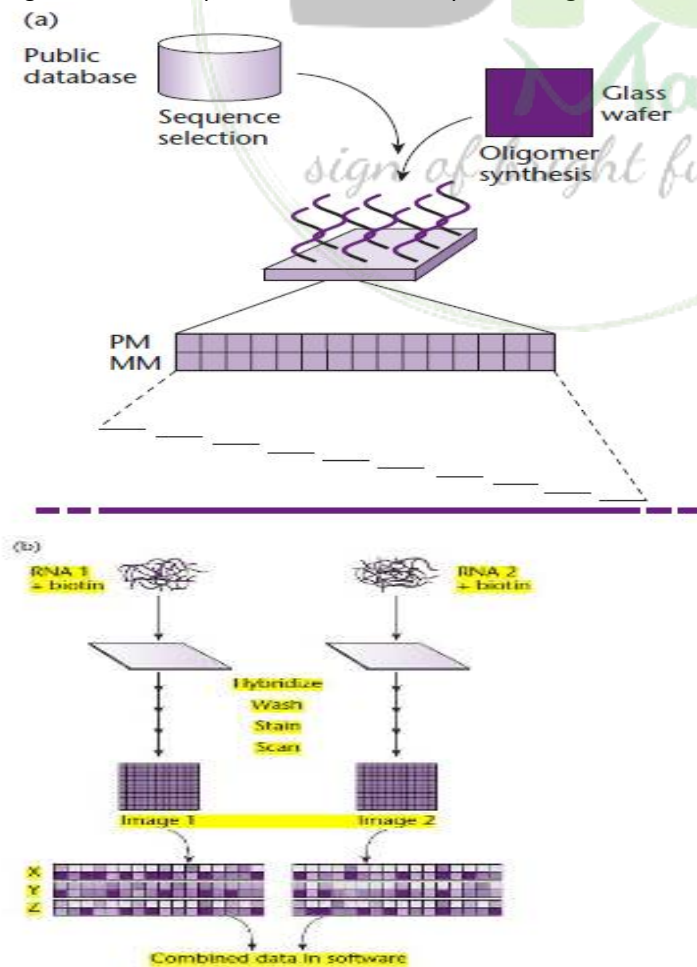
b)- Binding of DNA molecules to epoxy-derivatized surfaces

Spotted DNA arrays

Spotted DNA arrays are produced by printing DNA samples on treated microscope slides. It is made by transferring or spotting DNA clones or PCR product individually onto a solid support where they are immobilized

2010 Oligonucleotide chips

Oligonucleotide chips are manufactured by in situ oligonucleotide synthesis



Principle of chip hybridization

202 Microbial gene expression analysis

The predominate application of DNA microarray has been to measure gene expression levels

203 Profiling in human disease

Arrays have been used to investigate transcriptional profiles associated with human disease and to identify novel disease markers and potential new drug targets

Genotyping

Microarrays have been widely used as single nucleotide polymorphism (SNP) genotyping platforms

204 Phage display

Phage display is the technology that allows expression of exogenous polypeptides on the surface of phage particles

205 Screening phage display libraries

The most common screening method is based on enriching the phage clones with binding affinity for the target by a process called biopanning

206 Applications of phage display

Applications of phage display includes determination of protein to protein interactions, to determine enzyme specificity and to generate target specific antibodies

207 Knock outs

A gene knock out is a genetic technique in which one of an organism's gene is made inoperative that is simply called as knocked out

In molecular cloning, a knock-in refers to a gene manipulation method that involves the insertion of a protein coding cDNA sequence at a particular locus in an organism's chromosome

208 siRNA technology

Small interfering RNA (siRNA) is the most commonly used RNA interference (RNAi) tool for inducing short-term silencing of protein coding genes. siRNA is a synthetic RNA duplex designed to specifically target a particular mRNA for degradation

209 Applications of siRNA

siRNA) is the method of choice to target specific genes for silencing and has provided immense potential as therapeutic tools