

VU LEARNING HUB

vulearninghub.xyz

HIGHLIGHTED HANDOUT

BT511T

Genomics and Proteomics

BT511T by VU Learning HUB

Prepared for VU students • Free study resources

<https://vulearninghub.xyz>



Virtual University of Pakistan

FEDERAL GOVERNMENT UNIVERSITY

COURSE HANDOUTS

Introduction to Biotechnology

BT511T

Order Your VU Books Now

Get all Virtual University books in high-quality hard copy, delivered right to your doorstep. Visit vubookshop.pk



Order Now

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.

Branches of Biotechnology:

(i) White Biotechnology:

Development of processes and microorganisms for Industrial processes.

Key concept:

Scale of operation.

Low volume high price.

Differentiate between low volume high price products and high volume low price products Proteases (blue), lipases, amylases, cellulases.

White Biotechnology Example:

Enzyme production

(ii) Red Biotechnology:

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

(iii) Green Biotechnology:

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

An additional gene was inserted into the normal genome of this corn that triggers the production of proteins that then protect the plant from harmful insects. This reduces the need for pesticides and increases crop yields

Green Biotechnology Example:

Bt corn from Syngenta

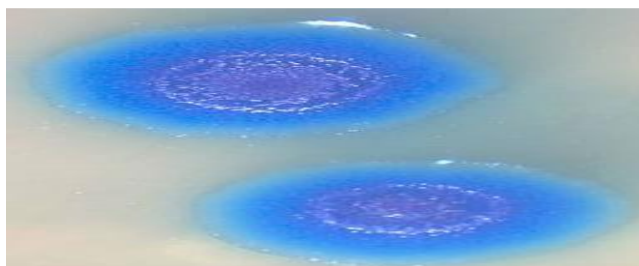
Bt cotton from Monsanto

2.Modern Biotechnology: (Principles & Applications)

Biotechnology in the 21st Century:

Ultimate reboot cells:

These are synthetic cells. We can insert DNA of our own choice in these cells and can determine cell type ultimately. That's why they are called ultimate reboot cells.



Historically it has been more about bacterial engineering but now it has shifted to eukaryotic engineering. (Craig Venter's programmable cells)

Growing Teeth:

Taking gum cells and ultimately converting them into teeth.



Gum cells directed to grow teeth.

Work of Prof. Paul Sharpe at King's College London.

Biological tooth replacement - a step closer

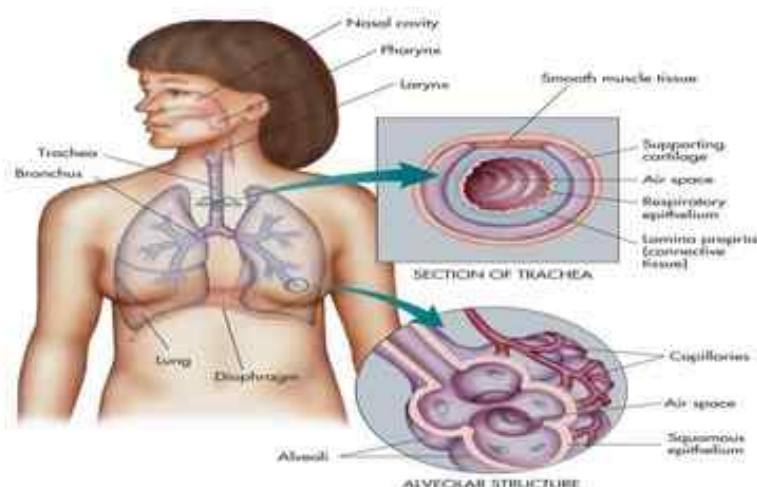
New research published (Posted on 11/03/2013) in the *Journal of Dental Research* describes an advance in efforts to develop a method to replace missing teeth with new bioengineered teeth generated from a person's own gum cells. The research is led by Professor Paul Sharpe, an expert in craniofacial development and stem cell biology at King's College London's Dental Institute, and was funded by the National Institute for Health Research (NIHR) Biomedical Research Centre at Guy's and St Thomas' NHS Foundation Trust and King's College London, UK.

Current implant-based methods of whole tooth replacement fail to reproduce a natural root structure and as a consequence of the friction from eating and other jaw movement, loss of jaw bone can occur around the implant.

Professor Sharpe concluded: 'Epithelial cells derived from adult human gum tissue are capable of responding to tooth inducing signals from embryonic tooth mesenchyme in an appropriate way to contribute to tooth crown and root formation and give rise to relevant differentiated cell types, following in vitro culture. These easily accessible epithelial cells are thus a realistic source for consideration in human biotooth formation. The next major challenge is to identify a way to culture adult human mesenchymal cells to be tooth-inducing, as at the moment we can only make embryonic mesenchymal cells do this.

Claudia's Trachea:

Spanish doctors took a piece of trachea (7 cm) from a deceased donor and they got rid of donor cells, ultimately it was just a bear cartilage, they pushed bone marrow cells (taken from patient named Claudia) and transformed them into cartilage cells. Ultimately a trachea was formed.



Colombian woman named Claudia Costillo was the first to get her trachea replaced which was made of her own cells.

Stem cells from her bone marrow were used and part of her wind pipe was used.

Claudia Castillo, the 30-year-old patient, had suffered a collapse of the tracheal branch of her windpipe leading to her left lung following a severe tuberculosis infection.

Left barely able to breathe, the decision was taken in March to attempt the windpipe reconstruction.

Spanish doctors started the process by taking a 7-centimetre section of windpipe from a deceased donor.

Researchers at the University of Padua, Italy, led by Maria Teresa Conconi, then used detergent and enzymes to purge the donated windpipe of all the donor's cells. After six weeks, all that was left was a solid scaffold of connective tissue.

Meanwhile, Birchall and his colleagues in Bristol took the stem cells from the patient's bone marrow and coaxed them in the lab into developing into the cartilage cells that normally coat windpipes.

Finally, the patient's cells were coated onto the donated tracheal scaffold over four days in a special bioreactor built at the Polytechnic of Milan in Italy.

Urinary bladder can be made out of our own cells with the help of stem cells.



Growing Blood Vessels:

We can take stem cells and can differentiate them into blood vessels.



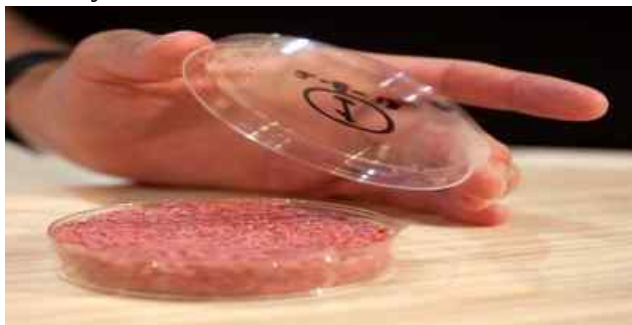
Chicken with an extra leg:

Chicken wings can be differentiated into chicken legs.



Growing Meat:

We can grow meat in laboratory.



3. Roots of Biotechnology:

6000 BC

Sumerians & Babylonians used **anaerobic respiration** of yeast to make beer.
(Sumerians are people from Southern Iraq).

4000 BC

Egyptians used **aerobic respiration** of yeast to leaven or rise bread and used mold to flavor cheese.
Fermented food is protected for longer and is easier to digest

1350 BC

Pregnancy test:

Barley grows – male child

Wheat grows – female child

A papyrus described a test in which a woman who might be pregnant could urinate on wheat and barley seeds over the course of several days: “If the barley grows, it means a male child. If the wheat grows, it means a female child. If both do not grow, she will not bear at all.” Testing of this theory in 1963 found

substance in the urine of pregnant women, and have speculated that elevated levels of estrogens in pregnant women's urine may have been the key to its success

400 BC

Use of salts to preserve food specially meat.

European Dark Period & Islamic Golden Era 1665 AD

First scientific best seller, written by Robert Hooke brought microscopy into scientific limelight

1676 AD Antonie Van Leeuwenhoek

Observation of microscopic single celled organisms.

Very fine lens maker, considered father of microbiology. He made around 200 microscopes and 500 lenses with various magnifications. His lenses were ultimately replicated by C. L Strong in 1957 almost after 300 years. 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" ☰

Scientists were pondering the nature of living things? What is the difference between living things and inanimate things? Theodore Schwann defined cell as a membrane bound structure containing nucleus. The body was not infused by a mysterious force, and life originated from the actions of cells. 1834 to

1839.....He was active as a scientist. 1834 - 1839 AD Cell Theory:

Theodore Schwann

“All living things are composed of cells and cell products”

Intellectual blind spot and 5 years of brilliance: Scientists were pondering the nature of living things? What is the difference between living things and inanimate things? Theodore Schwann defined cell as a membrane bound structure containing nucleus. The body was not infused by a mysterious force, and life originated from the actions of cells. 1834 to 1839.....he was active as a

Scientist. In that final year, he suffered a vicious personal attack by Fredrick Wohler. They mocked his idea that alcoholic fermentation was the result of yeast acting on sugars, insultingly drawing cartoons of yeast excreting wine through their wine bottled shaped imaginary asses. Prevailing theory at that time was that sugar transformed by reacting with air and nitrogenous substances in fruit juices. Most highly respected authorities denied Schwann any funds and his career was over.

1857 AD

Rudolf Virchow

“Omnis cellula e cellula”.

“Every cell arise from another cell”.

4_Roots of Biotechnology.

This laid the foundation of biotechnology.

1849

Industrial production of Baker's yeast.

1876

Kuhne creates the term “enzymes” and discovered trypsin.

Enzymes are biological catalysts.

1881

Industrial microbial production of lactic acid from lactic acid bacteria.

Teeth cavities are caused by lactic acid bacteria however lactic acid is widely used as a preservative in food industry. It is a very good descaler and soap scum remover.

1890

Development of first vaccines by Robert Koch and Louis

Pasteur. 1893

Industrial production of citric acid by Aspergillus Niger. Aspergillus niger is a mold.

Patent granted for enzymes in the detergent industry and leather production to Otto Rohm.

1917

Industrial fermentation process for Acetone, Butanol and Ethanol from starch.

(Scientist Chaim Weizmann)

Production of Cordite; Cordite is an explosive which was widely used in World War 1 and 2.

1928

Alexander Fleming discovers penicillin, *Penicillium Notatum*.

5.Discovery of DNA:

1869

Friedrich Miescher discovered nuclein. He found a substance inside nucleus and called it nuclein.

1878

Albrecht Kossel isolated non-protein component of nuclein & discovered nucleic acids & isolated 5 bases:

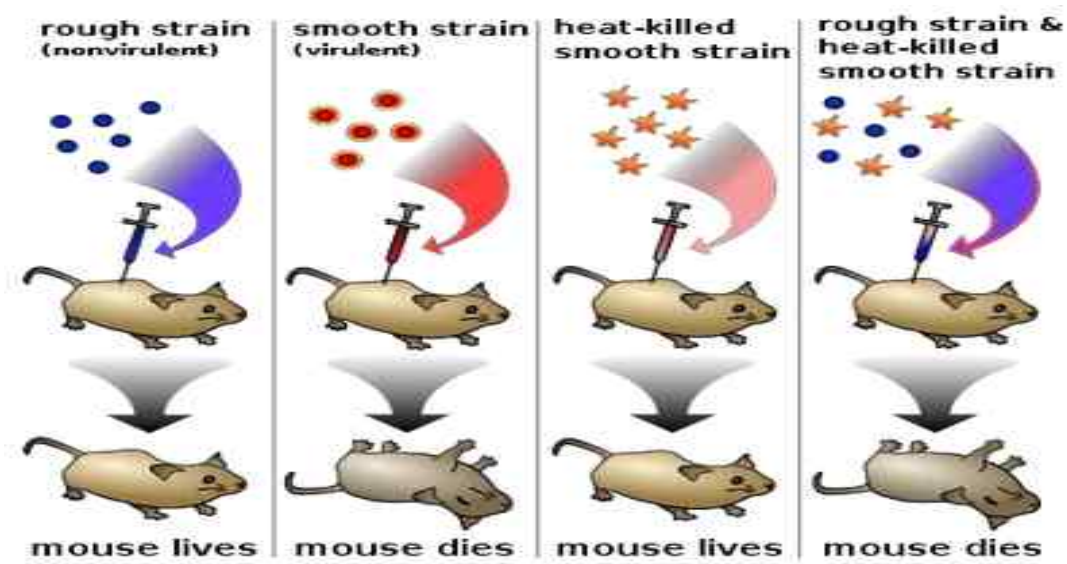
A, T, G, C & U.

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

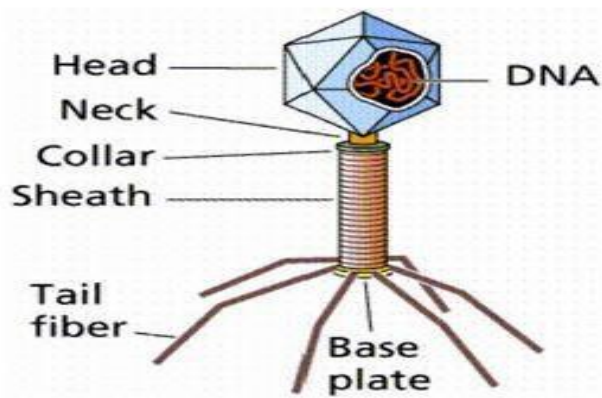


1

944 Avery, MacLeod & McCarty Experiment; They isolated DNA. 1952

Hershey-Chase Experiment; Radioactive phosphorus 32 and Sulfur

35, Sulfur is only present in protein.

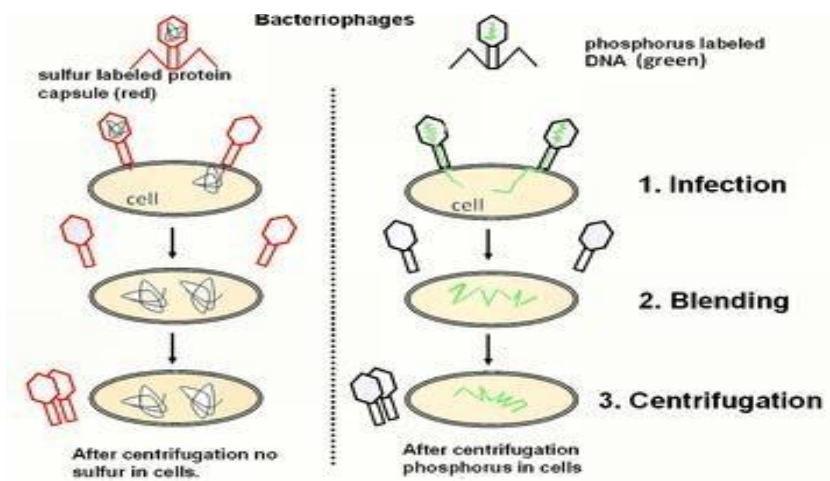


Rosalind Franklin (died at the age of 37) supervisor of Maurice Wilkins.

Wilkins said its alpha helical while Franklin denied.

Watson & Crick 1953

They proposed that DNA is a twisted ladder like structure. Sugar phosphate backbone is on outside while bases are inside.



6. Dawn of Biotechnology:

Spectrum of Eye Colour:

Once the structure of DNA was discovered, hunt for genes become red hot.

Curly hair:

Round hair follicles make straight hair, flattened or C-shaped hair follicles make curly hair, and oval hair follicles make wavy hair. Hair texture is a continuous trait, meaning that hair can be straight or curly or anywhere in between.

Multiple genes control hair texture, and different variations in these genes are found in different populations. For instance, curly hair is common in African populations, rare in Asian populations, and in-between in Europeans. Straight hair in Asians is mostly caused by variations in two genes—different genes from the ones that influence hair texture in Europeans.

If we discuss about hair genes then it is a continuous trait.

Tongue Rolling: Controlled by one gene but some environmental factors may play a role as people can learn to. > Point to think:

Can we control the traits?

Can we isolate the genes?

Can we edit the genes?

Can we introduce them in new species?

Freckles:

Freckles are controlled primarily by the single MC1R gene.

Restriction Endonucleases:

To introduce new genes we required something to cut the DNA at a specific location

Herbert Boyer, Stanley Norman Cohen and Paul Berg discovered Restriction endonucleases.

In 1973

- Boyer brought in restriction enzymes
- Cohen contributed with plasmid (pSC101)
- African Clawed Frog gene into a bacterium pSC101 had tetracycline resistance.

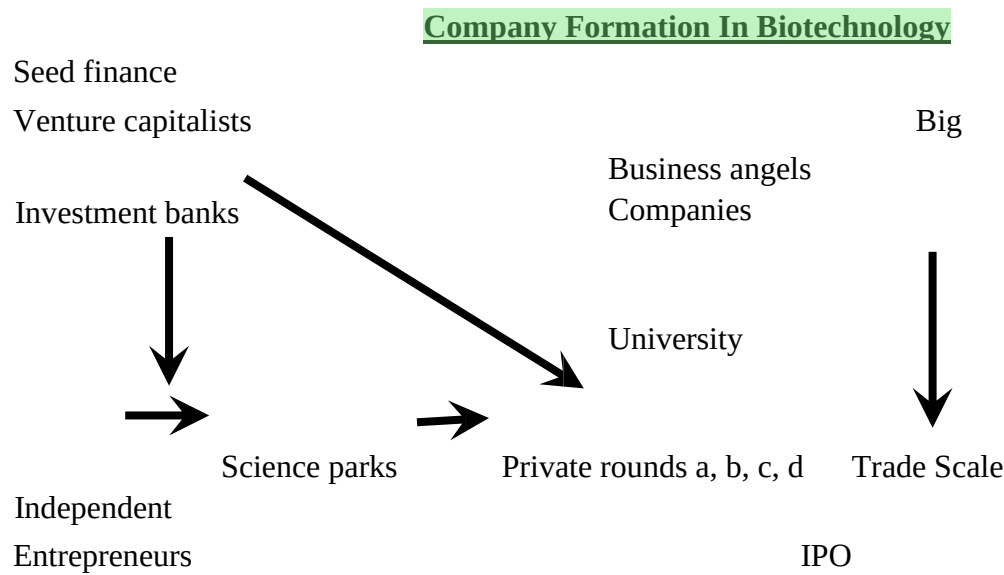
Negative Perception

“Genetic Engineering can give rise to something which can crawl out of the laboratory such as Frankenstein”, said by Local politician in Massachusetts

> Boyer went onto make Genentech in 1976. Robert Swanson was the first VC to invest in Biotechnology.

> Somatostatin was the first protein to be cloned in E. coli to produce a biopharmaceutical at large scale and Humulin was the first recombinant Insulin produced by Genentech but marketed by Eli Lilly.

In 1982, Amgen tendered an IPO raising \$42.3 million. This came on the heels of a remarkably successful IPO by rival Genentech, which created a short-lived biotech frenzy on Wall Street.



First generation of Biotech Companies:

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
Ribi 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
Repligen	04/86	\$17.5M

Commercializing Biotechnology

First generation of Biotech Companies

Company	Date of IPO	Amount Raised
OSI	04/86	\$13.8M
Cytogen	06/86	\$35.6M
Xoma	06/86	\$32M
Genzyme	06/86	\$28M
ImClone	06/86	\$32M
Genetics Institute	05/86	\$79M

Reference: from Alchemy to IPO (The business of Biotechnology) by Cynthia Robbins-Roth (2000)

8_ Biotechnology Giants post 2000



❖ DR. HERBERT BOYER

- Roche acquired Genentech for \$46.8 billion
- Current employees = 12,895
- 19 Breakthrough Biopharmaceutical Products



❖ DR. WINSTON SALSER

Applied Molecular Genetics

- Revenue \$20 billion
- Total Assets \$69 billion
- Net Income \$ 5 billion
- Current employees = 17,900
- 11 Breakthrough Biopharmaceutical Products



- Revenue \$9.7 billion
- Total Assets \$15 billion
- Net Income \$ 3 billion
- Current employees = 7,550
- 10 Breakthrough Biopharmaceutical Products
- ❖ Sir Kenneth Murray
- ❖ Charles Weissmann
- ❖ Phillip Sharp



- Revenue \$4.6 billion
- Net Income \$ 421 million
- Current employees = 12,000
- 16 innovative Biopharmaceutical Products
- ❖ DR. HENRY BLAIR was a research technician at Tufts University.

Gaucher’s Disease:

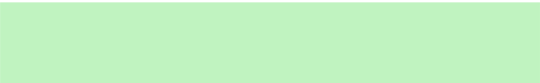
Genetic disease in which fatty substances (Sphingolipid named ceramide) accumulate in cells and certain organs. It happens due to the lack of an enzyme.

Molecular Reason:

Hereditary deficiency of the enzyme glucocerebrosidase.
Accumulation of a sphingolipid named ceramide that doesn’t get metabolised due to lack of

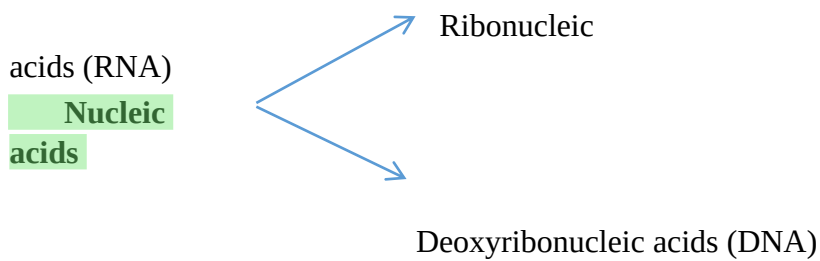
glucosidase enzyme. Symptoms:

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.



- Like proteins and polysaccharides, nucleic acids are also polymeric compounds.

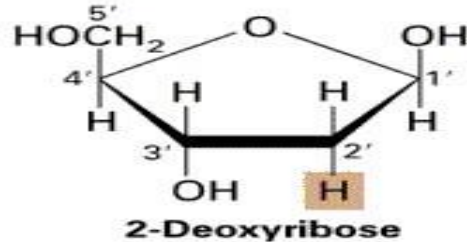
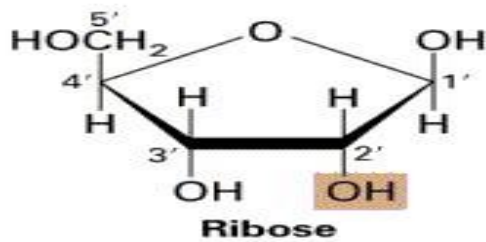
The repeating units in the nucleic acids are 'Nucleotides'.



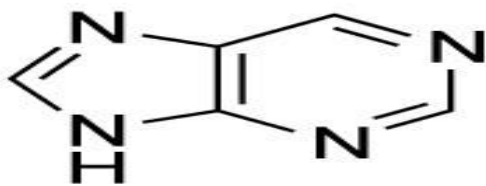
10. Chemical composition of DNA:

- DNA is a polymer of Deoxyribonucleotides.
- Deoxyribonucleotide is composed of three components:
 - Deoxyribose
 - Nitrogenous Base
 - Phosphoric acid

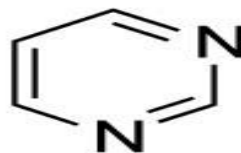
❖ Deoxyribose (a pentose sugar derivative):



❖ Nitrogenous Bases:

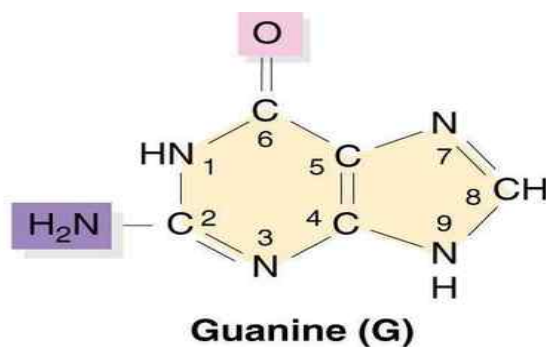
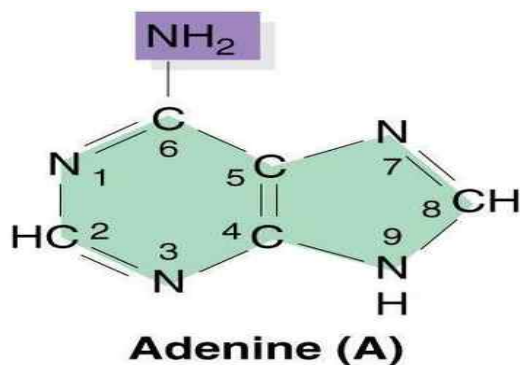


purine

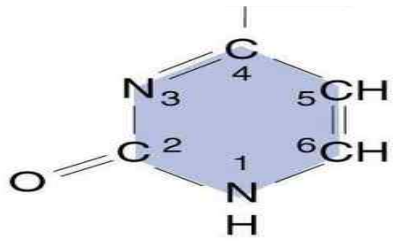


pyrimidine

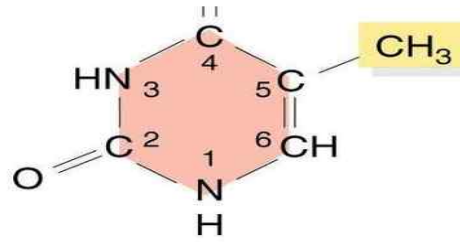
Purines:



Pyrimidines:

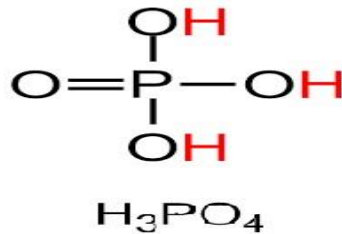


Cytosine (C)



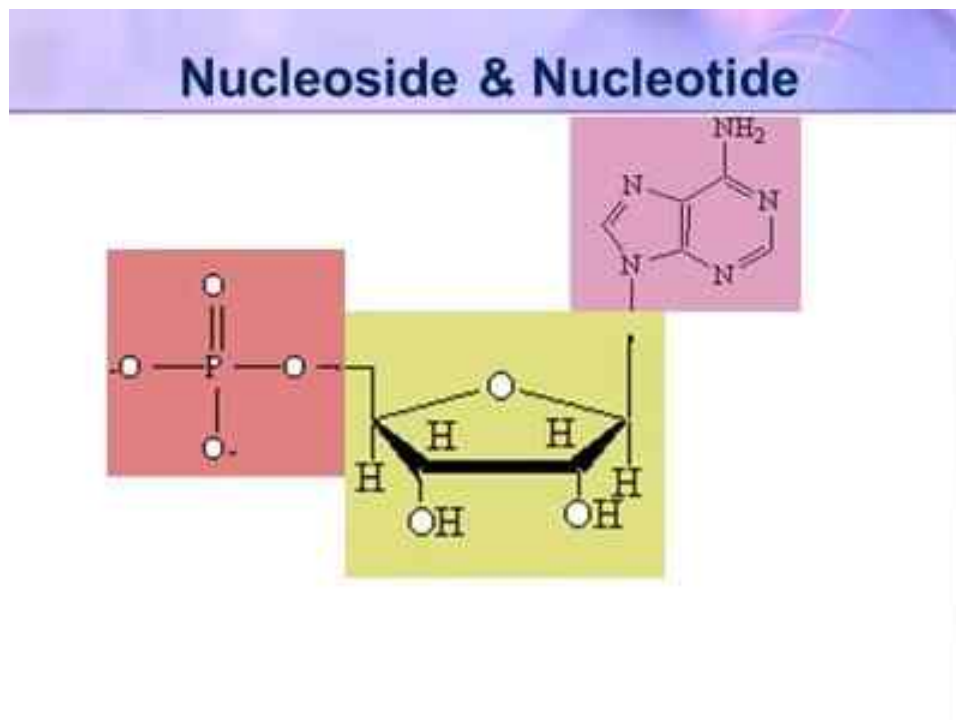
Thymine (T)

❖ Phosphoric acid:



11. Nucleoside & Nucleotide:

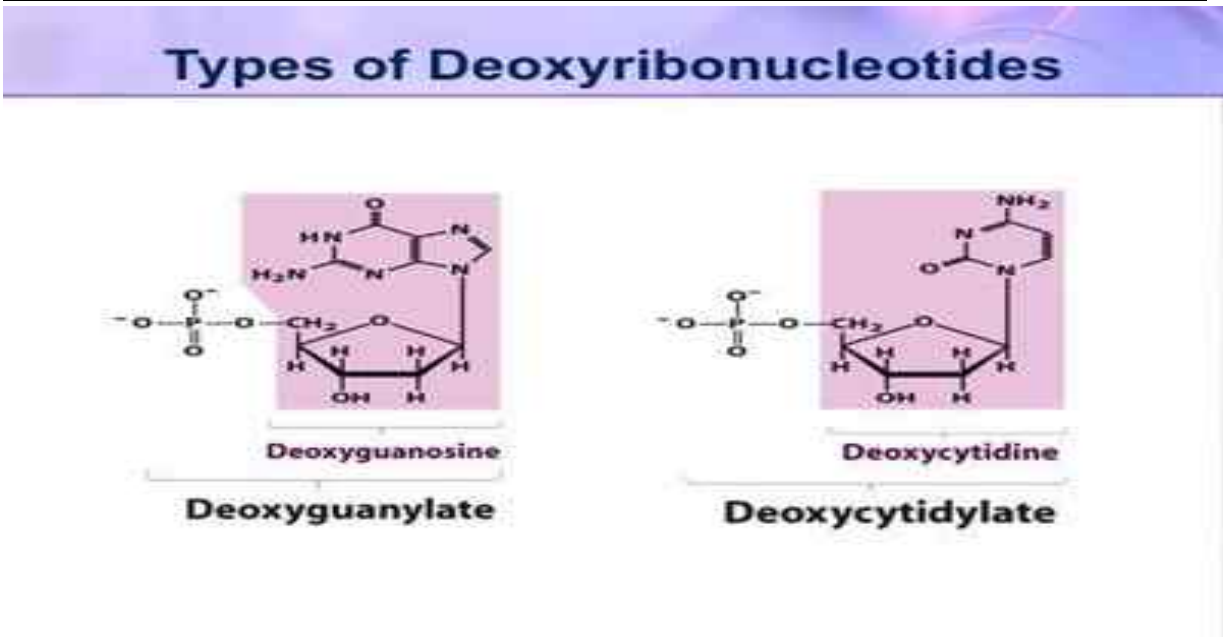
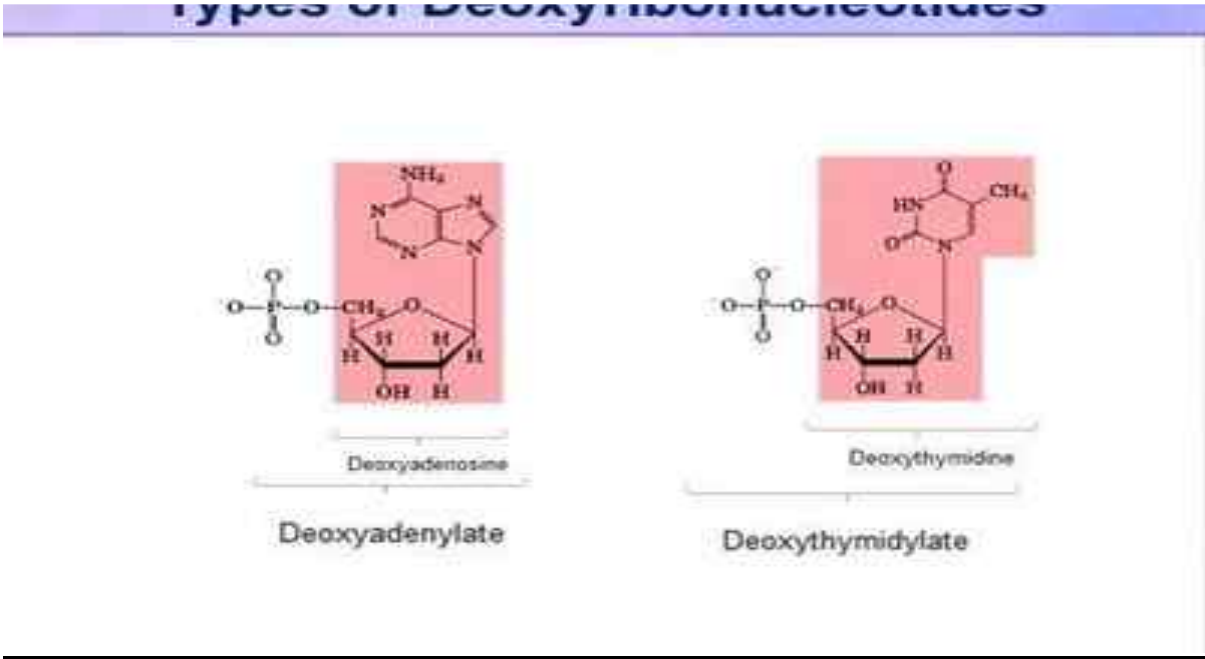
- A molecule containing all three components (Nitrogenous base, sugar molecule and phosphoric acid) is called a nucleotide.
- While a molecule without the phosphate group is called a nucleoside.



Nucleotide = Nucleoside +
Phosphoric acid Nucleoside =
Nucleotide – Phosphoric acid

12. Types of Deoxyribonucleotides:

- There are four types of Deoxy-ribonucleotides.



1. Deoxyadenosine (Deoxyadenylate)

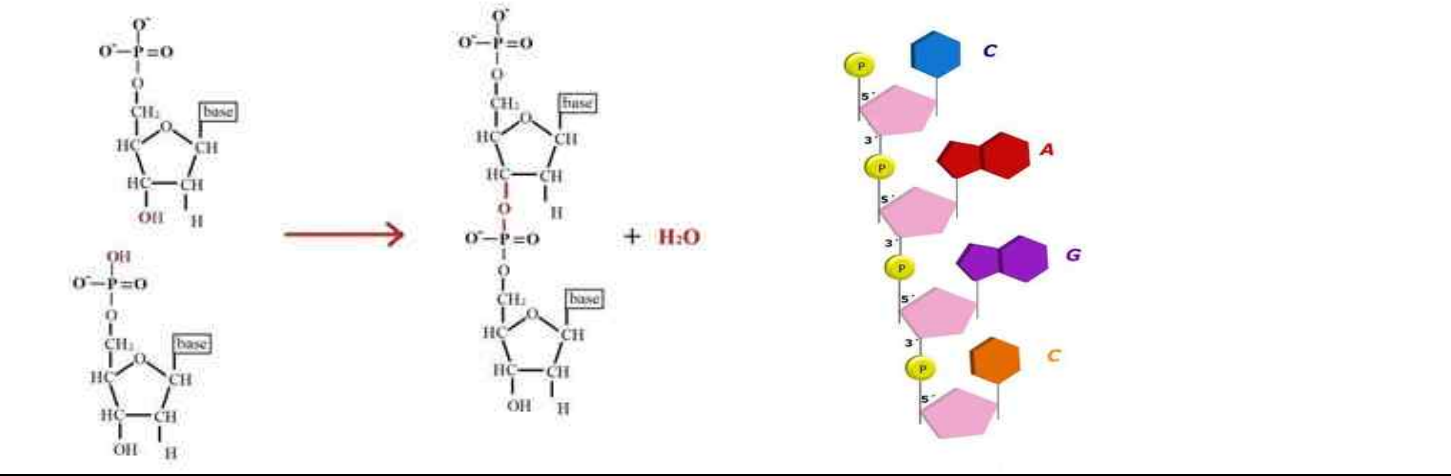
2. Deoxythymidine (Deoxythymidylate)

3. Deoxyguanosine (Deoxyguanylate)

4. Deoxycytidine (Deoxycytidylate)
- These four deoxy-ribonucleotides make the structural units of DNA.

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”.

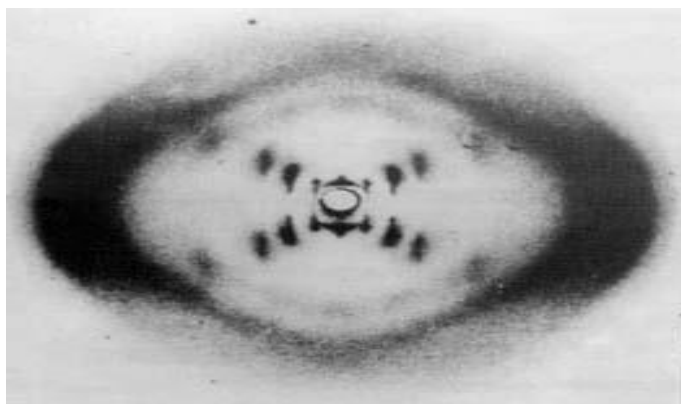
“Chargaff Rules”:

1. Base composition of DNA varies from one species to another.
2. The DNA isolated from different tissues of the same species has 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).

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



- From this pattern, they made two important findings:-

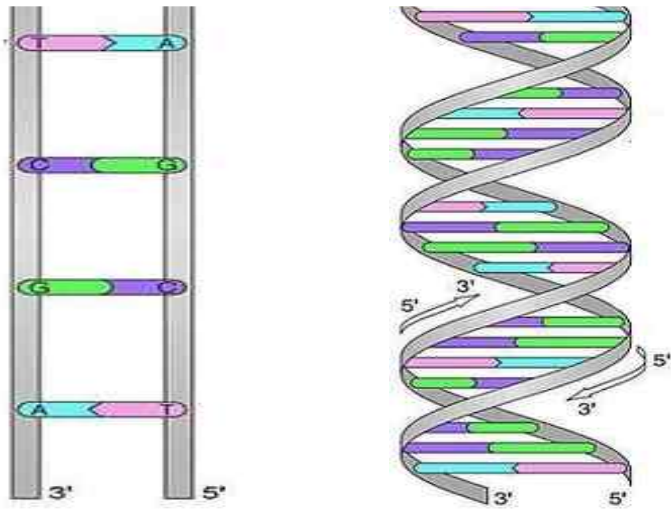
1. DNA molecules are helical.
2. The helices have two periodicities along their long axis, a primary one of 3.4 Å and a secondary one of 34 Å.

16_ Structure of DNA_2

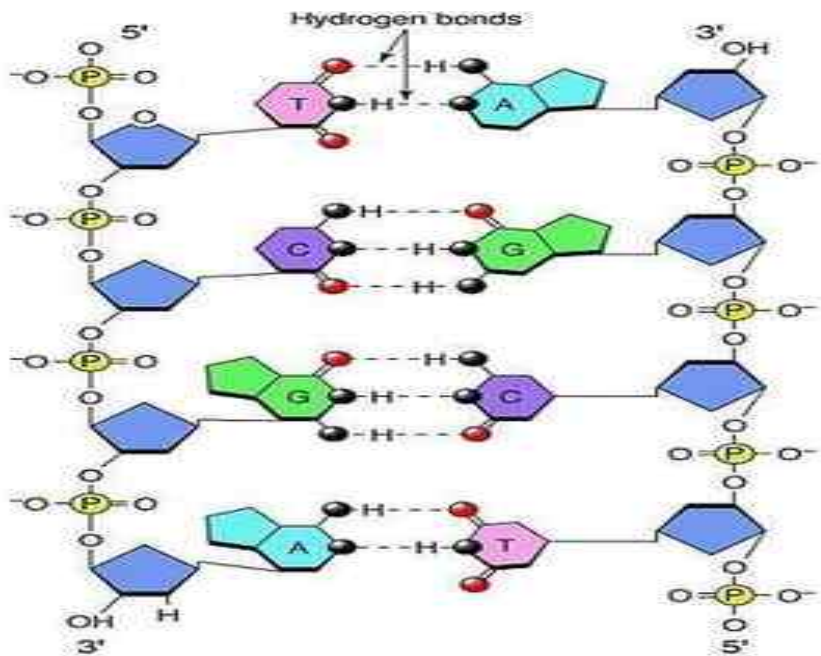
Work of Watson & Crick:

- James Watson and Francis Crick postulated a three dimensional model of DNA structure in 1953. They were awarded Nobel Prize in the same year, for their discovery. The major features of this model are as follow:-

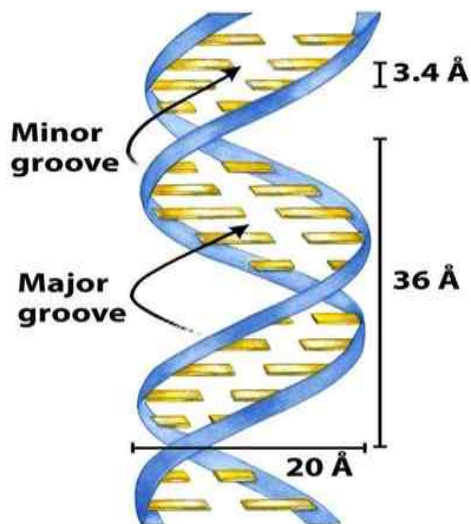
1. DNA consists of two helical polynucleotide strands which are wound around the same axis to form a right handed double helix.



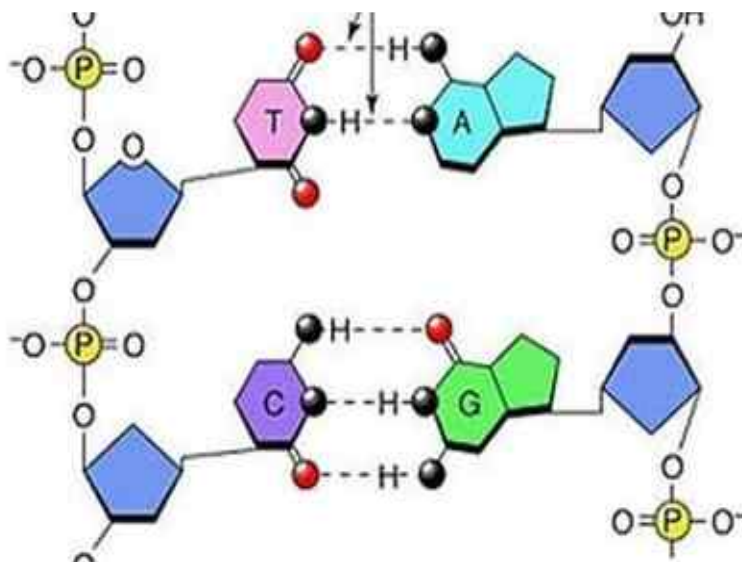
- 2. The hydrophilic backbones of alternating deoxyribose and phosphate groups are on the outside of the double helix, facing the surrounding water.
- 3. The nitrogenous bases of both strands are stacked inside the double helix lying perpendicular to the long axis of the helix.



- 4. The pairing and coiling of the two strands create a major groove and minor groove on the surface of the helix.



- 5. Each nucleotide base of one strand is paired in the same plane with a base of the other strand.

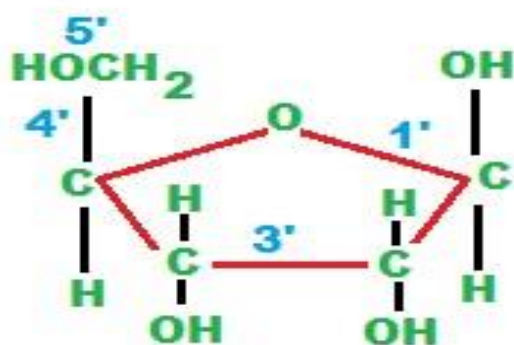


6. They also found that G pairs with C and A pairs with T due to the reason that they fit best within the structure.

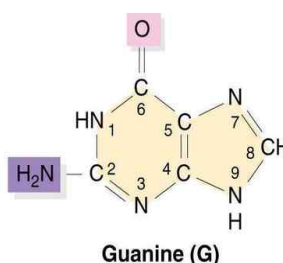
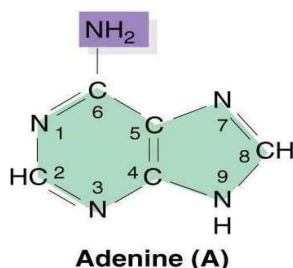
7. The two strands in DNA are present in antiparallel orientation i.e. their 5,3-phosphodiester bonds run in the opposite directions. 8. The two antiparallel strands of double-helical DNA are complementary to each other.

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 **Ribose (a pentose sugar):**



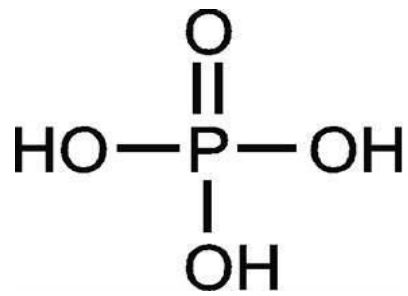
Nitrogenous Bases:



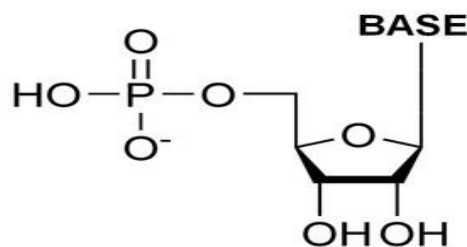
Phosphoric acid:

ribose).

Ideally it can make three phosphodiester linkages but normally in a poly nucleotide chain it makes two ester linkages.

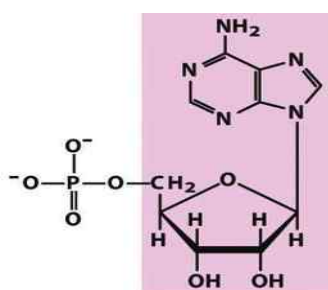


A Ribonucleotide:



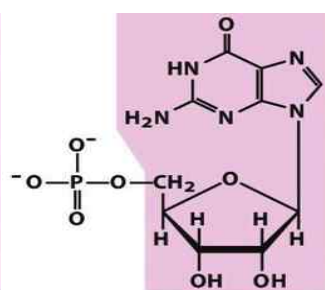
18. Types of Ribonucleotides:

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



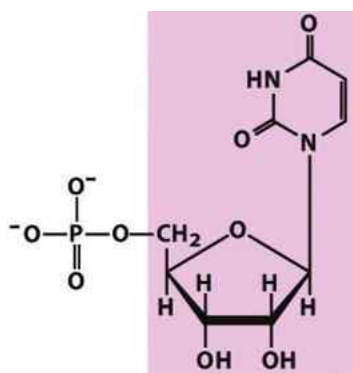
Adenylate (adenosine 5'-monophosphate)

Adenosine



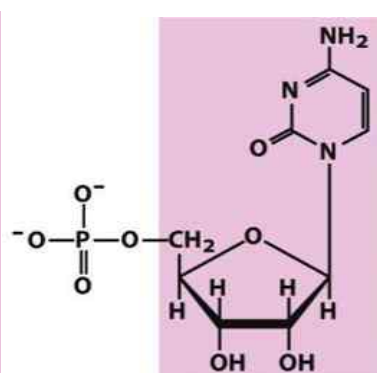
Guanylate (guanosine 5'-monophosphate)

Guanosine



Uridylate (uridine 5'-monophosphate)

**U, UMP
Uridine**

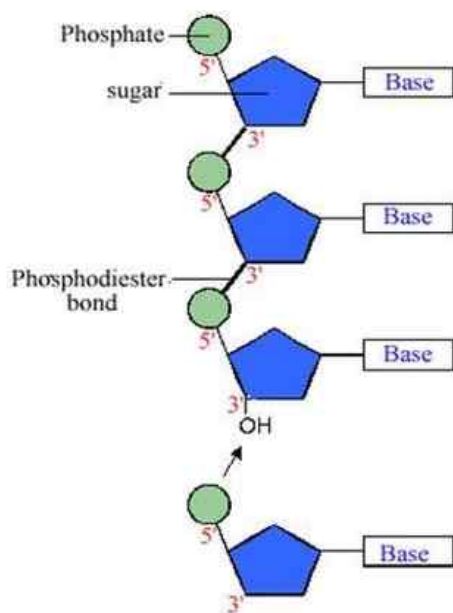


Cytidylate (cytidine 5'-monophosphate)

**C, CMP
Cytidine**



A Poly-Ribonucleotide:



(RNA molecules are much shorter.)

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)

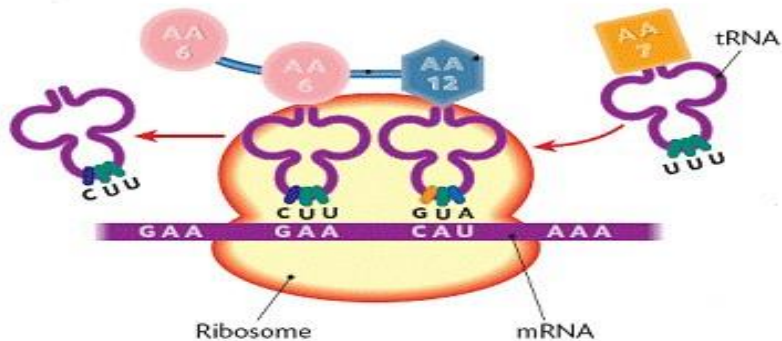
Messenger RNA (mRNA)

- 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.
- mRNA may be monocistronic (Coding one gene, i.e. it is responsible for the synthesis of only one type of protein) or polycistronic (contains information for more than one gene, i.e. it is responsible for the synthesis of more than one type of protein).
- The length of mRNA molecules is variable and it depends on the length of gene.

an amino acid. So it carries that amino acid to the site of protein synthesis that is ribosome. In this way tRNA acts as an adapter molecule.

The function of tRNA is to transfer amino acid from different sites of the cytoplasm to the site of protein synthesis in the cytoplasm.

- 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)

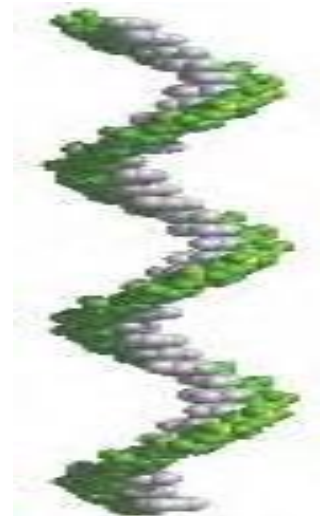
- Ribosomal RNAs are components of ribosomes.
- rRNA is a predominant material in the ribosomes constituting about 60% of its weight.
- It has a number of functions to perform in the ribosomes.

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.

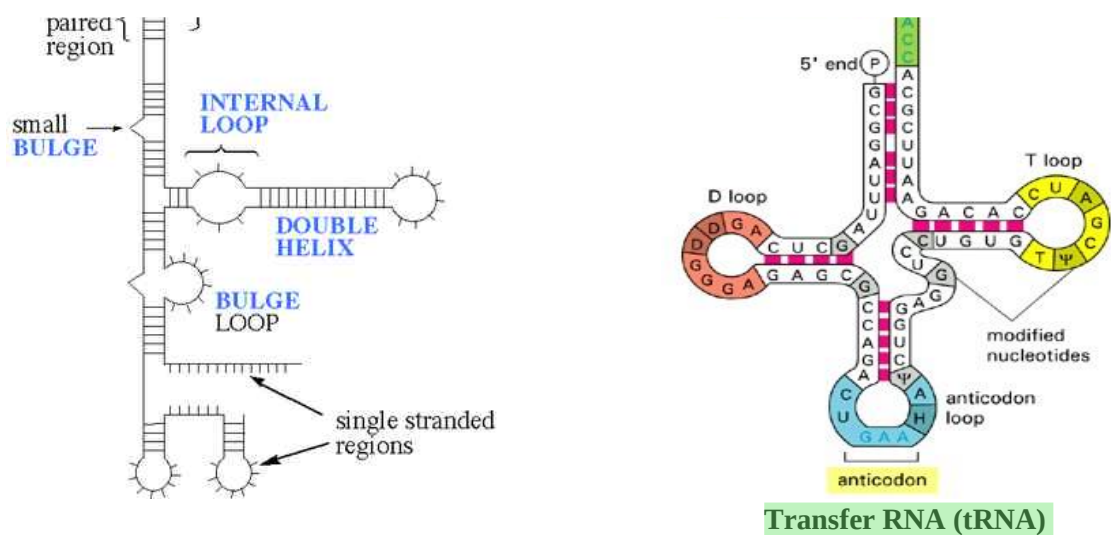
Base stacking interaction is stronger between purine-purine than between purine-pyrimidine or pyrimidinepyrimidine.

Messenger RNA (mRNA)



The bases are shown in gray, the phosphate atoms in yellow, and the riboses and phosphate oxygens in green.

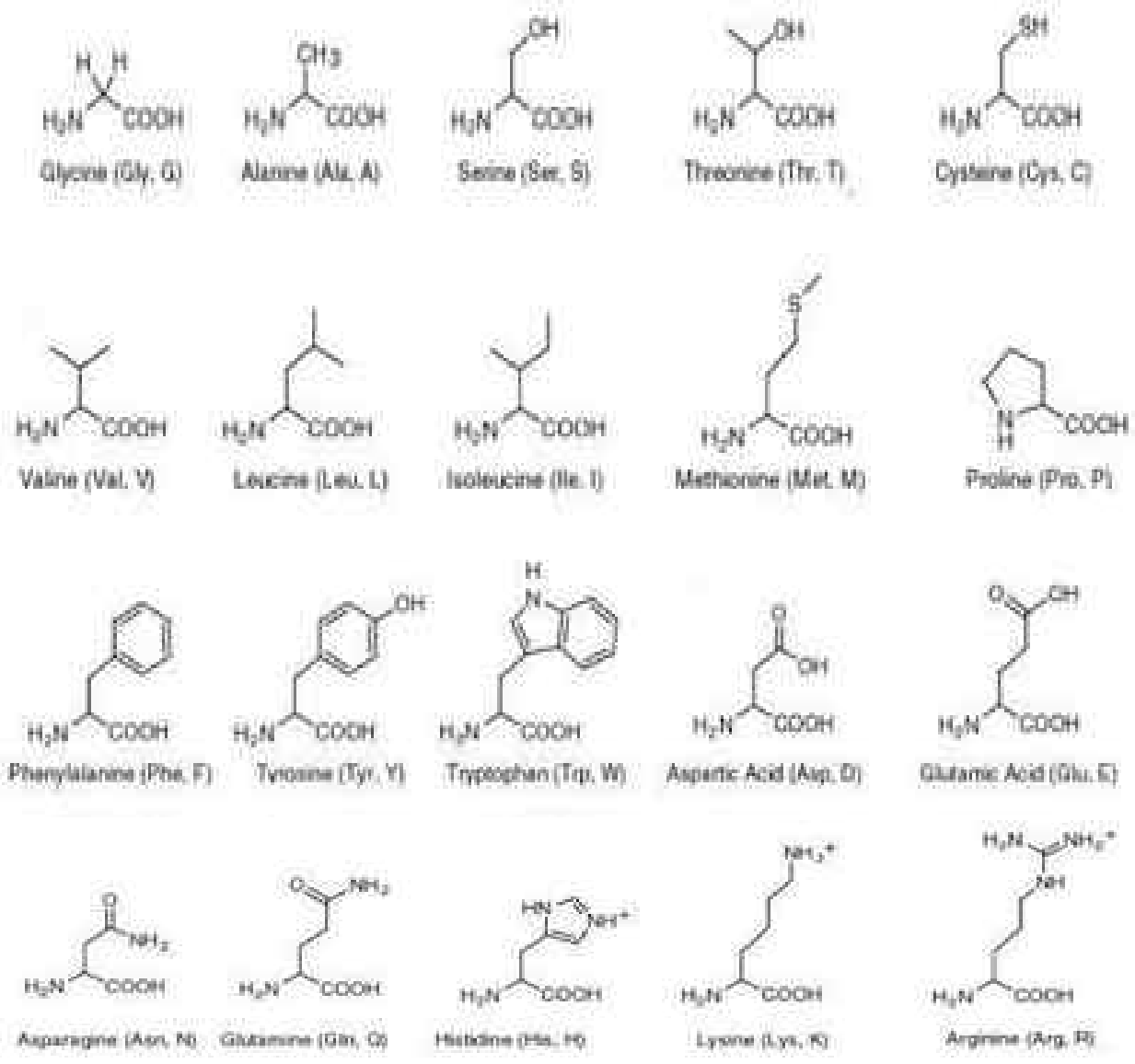
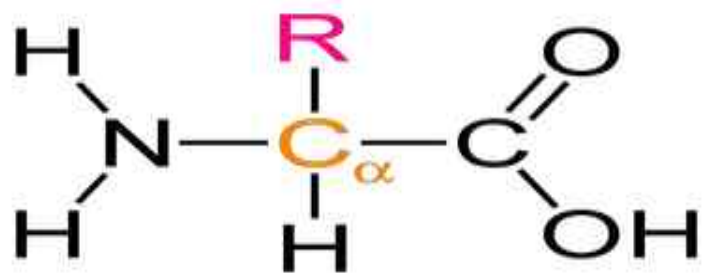
- Self-complementary sequences may occur in the RNA molecules which produce more complex structures.
- So RNA can base-pair with complementary regions of either RNA or DNA.
- RNA has no any regular secondary structure that serves as a reference point. The three-dimensional structures of many RNAs are 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.

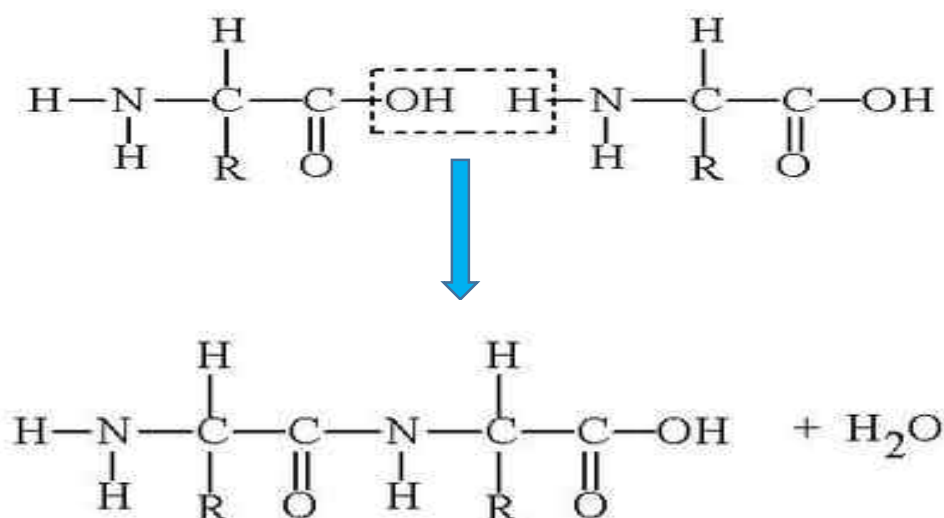


Transfer RNA (tRNA)

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 amino acids.





- Amino acids can successively join to form dipeptides, tripeptides, tetrapeptides, oligo peptides and polypeptides.

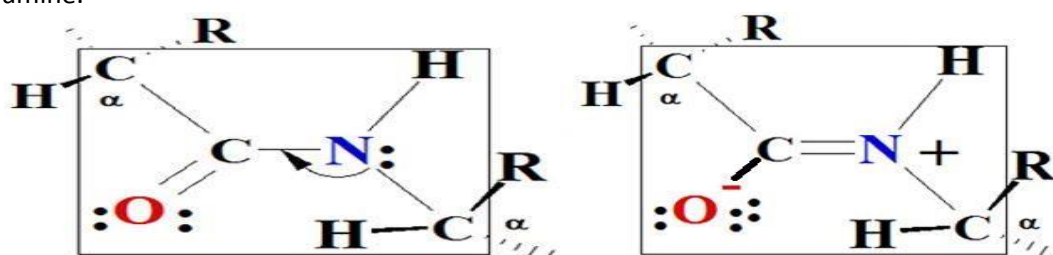


22. Primary structure of proteins:

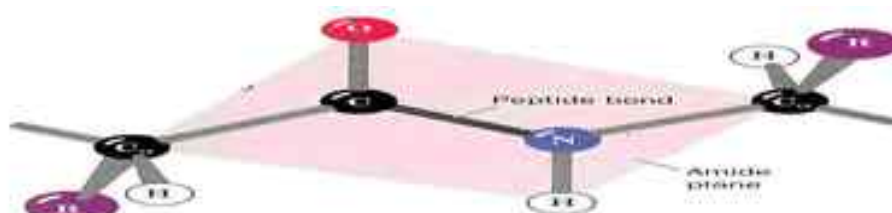
- Primary structure or covalent structure of protein refers to the amino acid sequence of its polypeptide chain.
- Each type of protein has a unique amino acid sequence.
Proteins are different from one another on the basis of their amino acid sequence.

Peptide Bond Is Rigid and Planar

- 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:-

1. α -helix

A polypeptide chain is helically coiled around an imaginary axis and it makes α -helix. It is a prominent structure in the proteins.



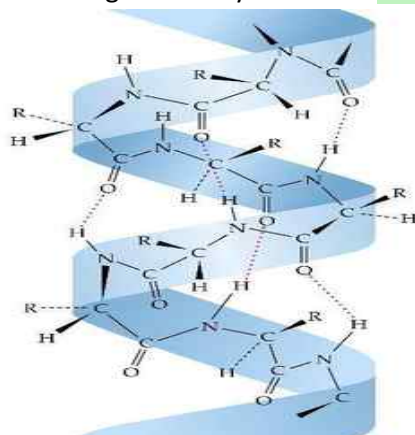
2. β - conformations

Polypeptide chains stretch side by side longitudinally and form sheets. These sheets are called β - sheets.

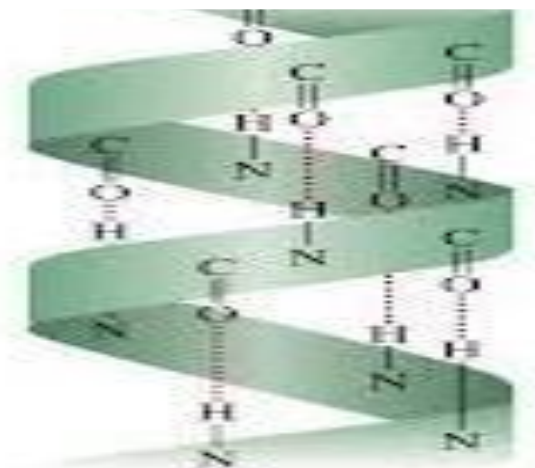


24. α - Helix:

- 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 helix have conformations with $\psi = -45$ to -50 and $\phi = -60$.

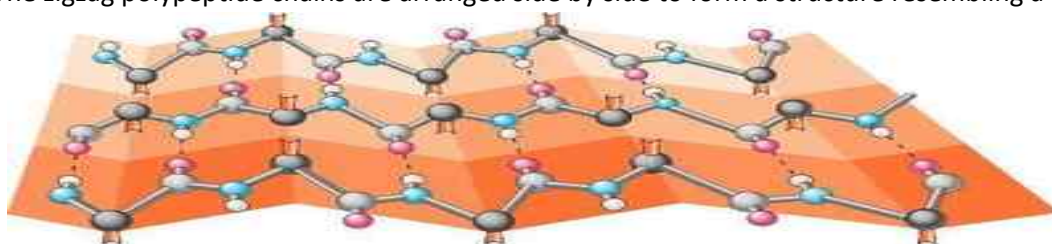


The structure is stabilized by a hydrogen bond between the hydrogen atom attached to the electronegative nitrogen atom of a peptide linkage and the electronegative carbonyl oxygen atom of the fourth amino acid on the amino-terminal side of that peptide bond.

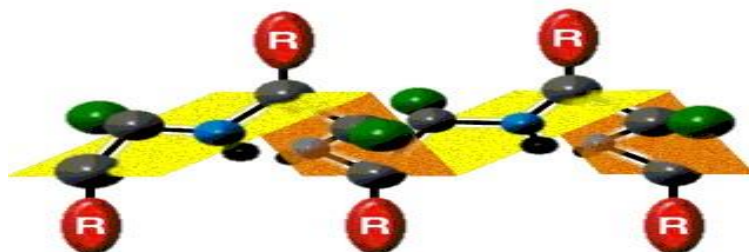
- About one-fourth of all amino acid residues in polypeptides are found in α -helices while in some proteins it is the predominant structure.

25. β - Pleated Sheets:

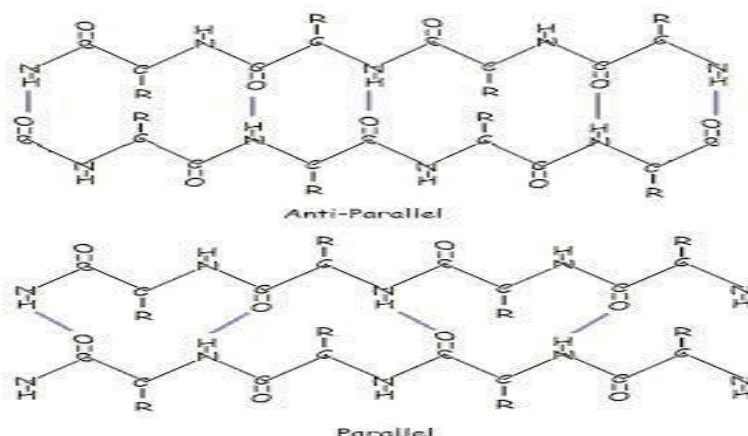
- Pauling and Corey predicted a second type of secondary structure which they called **β -sheets**.
- This is a more extended conformation of polypeptide chains.
- The backbone of the polypeptide chain is extended into a zigzag structure.
- The zigzag polypeptide chains are arranged side by side to form 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 sheet can be either parallel or antiparallel.



Difference b/w Hydrogen Bonding in Parallel

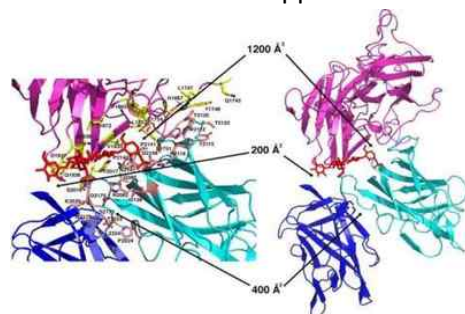
& Anti-parallel β -pleated sheets:

The hydrogen bonds are comparatively straight in case of antiparallel β -pleated sheets. While hydrogen bonds are slightly distorted in case of parallel β -pleated sheets. That's why anti-parallel β -pleated sheets are more stable as compared to parallel β -pleated sheets.

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

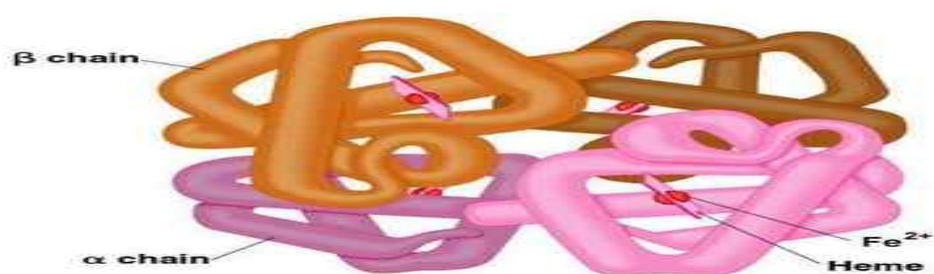


- It includes longer-range aspects of amino acid sequence.
- Amino acids that are far apart in the polypeptide chain may interact within the completely folded structure of a protein.
- Interacting segments of polypeptide chains are held in their characteristic tertiary positions by different kinds of 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.

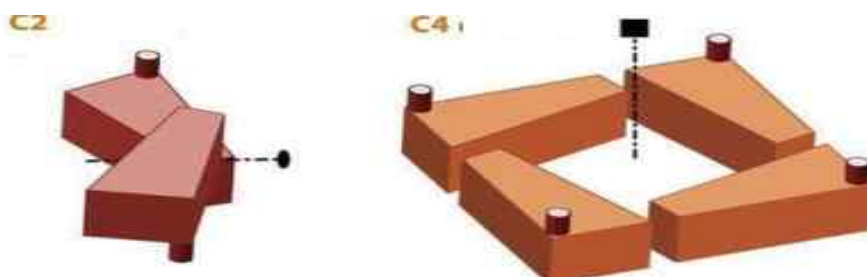


27. Quaternary Structure of Proteins:

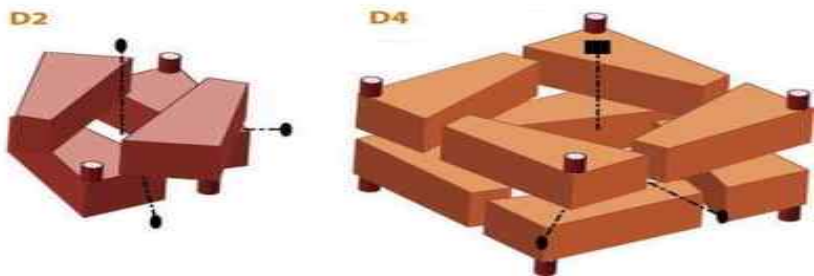
- Some proteins contain two or more separate polypeptide chains or subunits, which may be identical or different.
- The spatial arrangement of these subunits is known as a protein's quaternary structure.
- A multi-subunit protein is also referred to as a multimer.
- A multimer with just a few subunits 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 pattern.
- Oligomers can have either rotational symmetry or helical symmetry.
- There are several forms of rotational symmetry. The simplest is cyclic symmetry, involving rotation about a single axis.



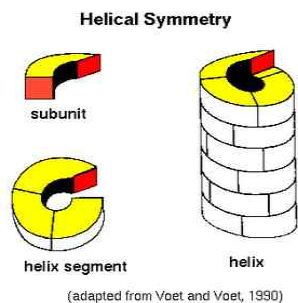
present.



- More complex rotational symmetries include icosahedral symmetry. An icosahedron is a regular 12-cornered polyhedron having 20 triangular faces.

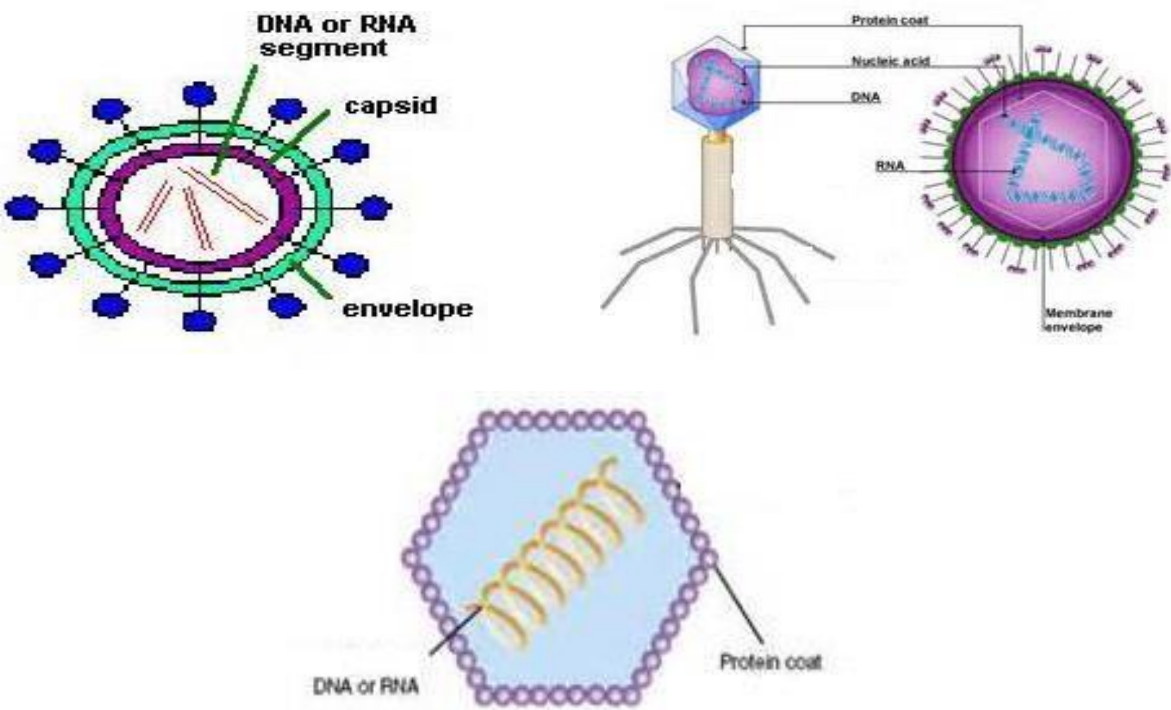


- The other major type of symmetry found in oligomers is helical symmetry.

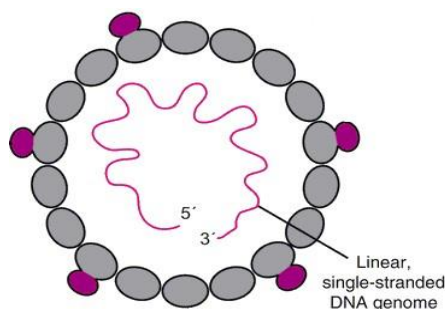
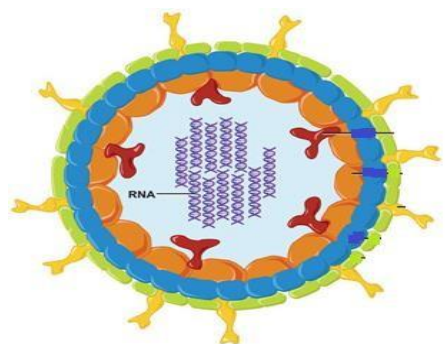


28.Genetic Materials in Viruses:

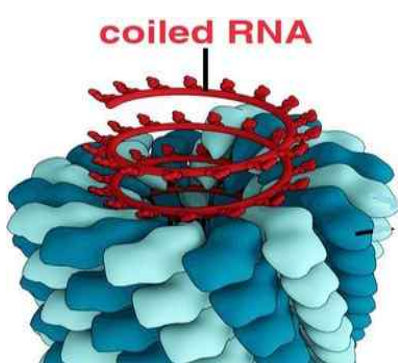
Viruses are exceptionally simple and extremely small microorganisms. They have a very simple structural organization consisting of a molecule of nucleic acid and a protein coat.



- 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.



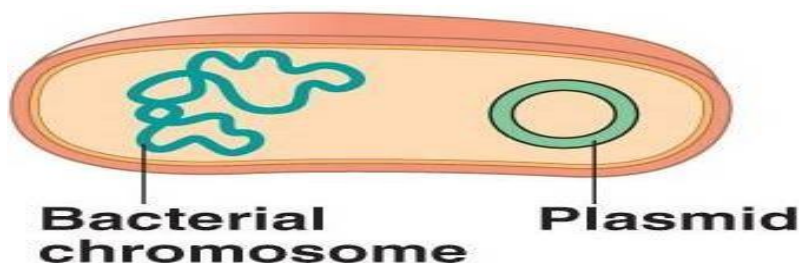
- The viral nucleic acid can be linear or circular.
- Some viruses may contain coiled RNA. • In some viruses, the nucleic acid is in several separate segments.



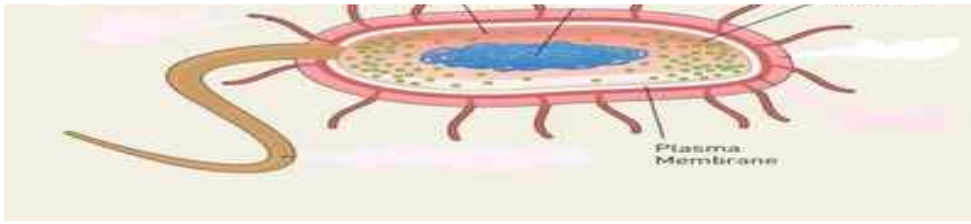
- The percentage of nucleic acid in relation to protein is about 1% for the influenza virus and about 50% for some bacteriophages.
- The total amount of nucleic acid varies from a few thousand nucleotides to as many as 250,000 nucleotides. *E. coli*'s chromosome consists of approx. 4 million nucleotide pairs.

29.30. 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 very long (up to 1mm).
- It is looped and folded and attached at one or several points to the plasma membrane.

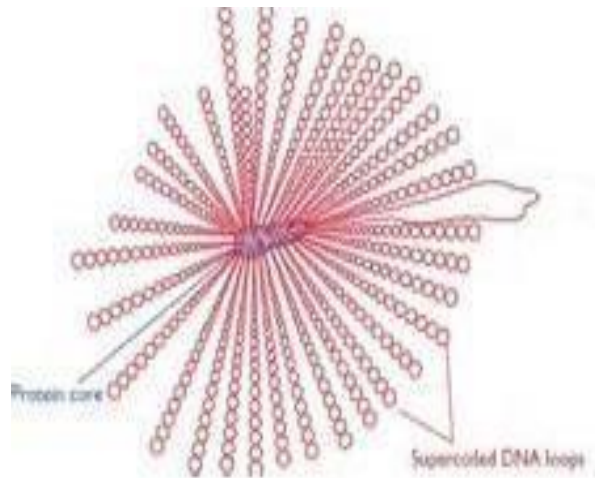
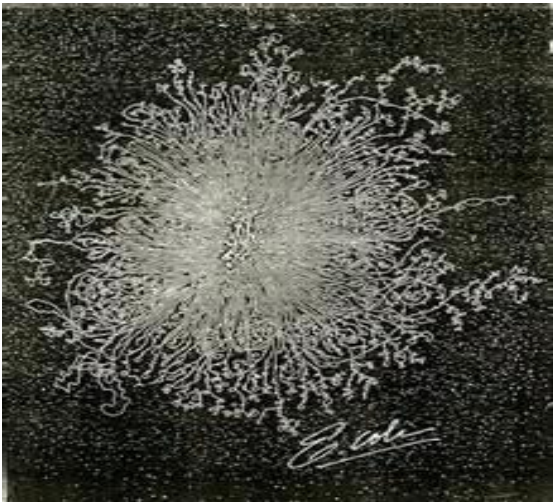


- Specific proteins interact with the bacterial DNA to form a highly condensed nucleoprotein complex called the **nucleoid**.

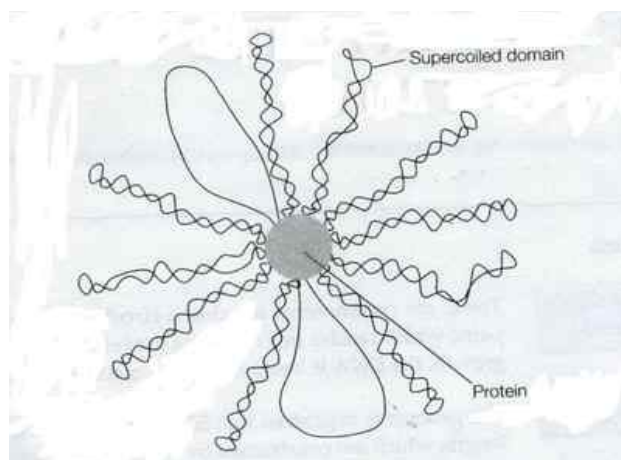


It is highly compact and occupies about a quarter of the cell's volume.

- 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.



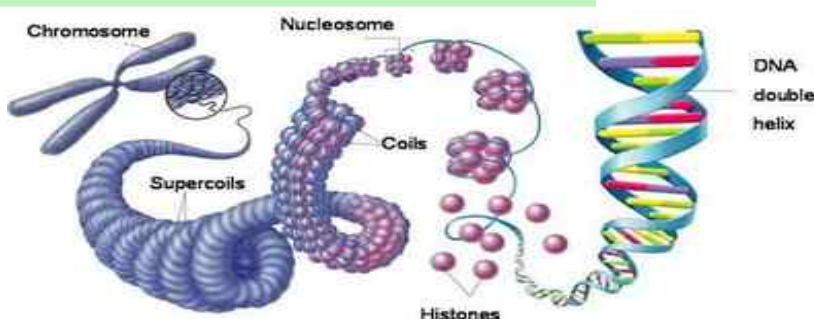
- So a nick that causes one super-coiled loop to relax would have no effect on other super-coiled loops.
- The super-coiled loops are dynamic structures which change during cell growth & division.
- An *E. coli* chromosome is estimated to have about 400 super-coiled loops.
- Each loop has an average 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.

31. Organization of Genetic Material in Eukaryotes:

- The genetic material (DNA) of eukaryotic organisms is organized in the form of chromosomes.
- The chromosomes of eukaryotic cells are larger and more complex than those of prokaryotes.



- Each un-replicated chromosome consists of a single molecule of DNA. While if this chromosome is replicated then two molecules of DNA will be present in it.
- If stretched out, some human chromosomes would be several centimetres long.
- To package such a tremendous length of DNA into this small volume, each DNA molecule is coiled again and again and tightly packed around histone proteins.
- As eukaryotic chromosomes are not circular, so instead of super-coiling, the mechanism of packaging involves winding the DNA around special proteins, the **histones**.
- **DNA with bound histones in the eukaryotes is called as chromatin.**
- Chromatin consists of roughly spherical subunits, the **nucleosomes**, each containing approx. 200 bp of DNA and nine histones.
- A condensed mitotic chromosome is about 50,000 times shorter than fully extended DNA.
- **Highly condensed chromatin is known as heterochromatin.**
- **The more extended form is known as euchromatin.**



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.

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.

Conclusion:

- 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:

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 and the study of proteome is called as proteomics.
- 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.

- 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

Novel Therapeutics

- 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. • issues within different organisms.

35. Genomes Characterization –Techniques

Major Techniques used for Genomes Characterization:

- Cloning
- Hybridization
- PCR amplification
- Sequencing
- Computational tool

Genomes Characterization Techniques - Cloning

- Genomes digested with restriction enzymes and inserted in vectors to produce genomic libraries.
- Examples of such libraries are:
B
A
Cs
Y
A
Cs
- Bacterial Artificial Chromosomes (BACs) were used to prepare genomic libraries.
- Yeast Artificial Chromosomes (YACs) were used to prepare genomic libraries for the sequencing of the genomes.

Genomes Characterization Techniques - Hybridization

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

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 (Figure):

CTACAGTGGCTGCTCTCTTCTG GTTATGTTGCTGGACAT



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

Once the genome of any organism is sequenced there are certain steps to analyze the sequenced data to get information from the sequenced data of the genomes.

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

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

- Help us to 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.

Genomes Research – Agriculture, Livestock Breeding, and Bioprocessing

- Develop bio-pesticides.
- 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. Genes and Size of Genomes:

Size of Genomes

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

- Genomes of eukaryotes are usually larger
- Most plants and animals have genomes greater than 100 Mb.
- Humans have genome size of 3,000 Mb

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

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
- 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.

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

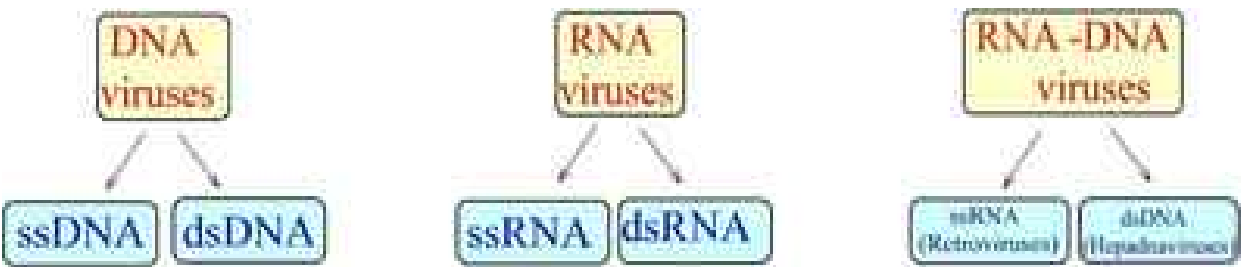
- Often a circular genome
- lambda = 48,502 bp in size

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.



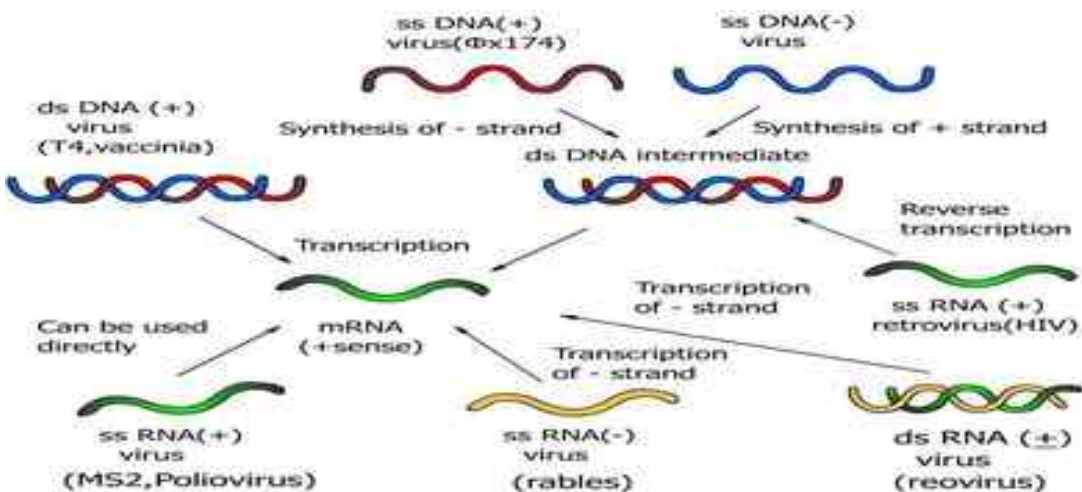
Viruses and Number of Genes:

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
Qβ	<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

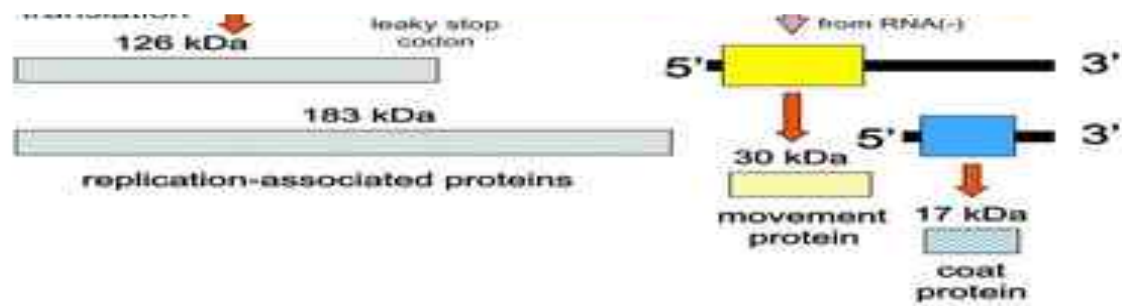
Genomes in Virions:

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.



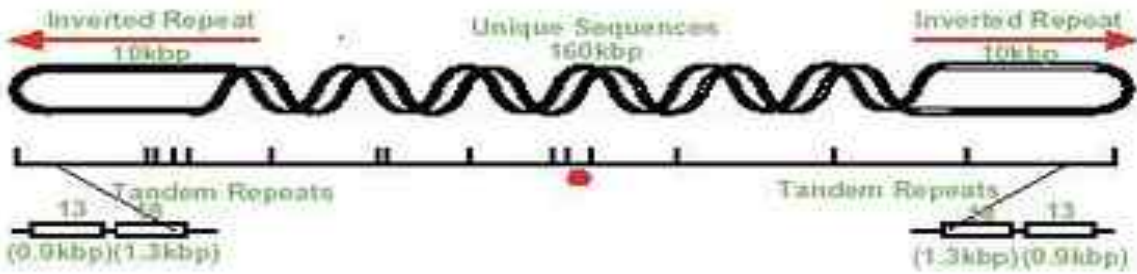
Genome of Tobacco Mosaic Virus: • Single, 6400 nucleotides RNA, 3

Essential Genes



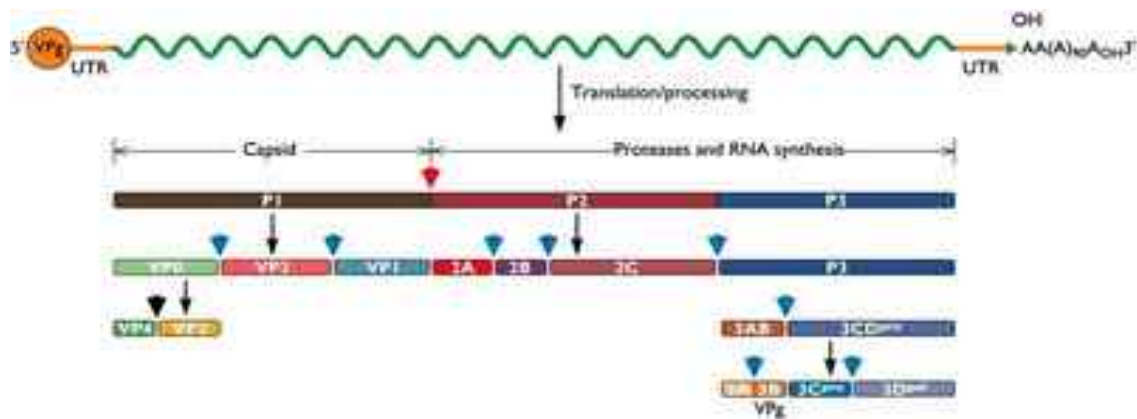
Genome of Poxvirus – A typical large dsDNA Virus

- 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 – Conclusion

- Viral genomes can be
 - ssRNA
 - dsRNA
 - ssDNA
 - dsDNA
 - Linear
 - Circular

40.Bacterial Genomes:

(Genomes of Bacteria)

- Small organisms carry high coding density (85-90%) i.e. 85-90% of their genome is coding in nature.
- 1 gene per 1000 bases in prokaryotes
- Large variation in genome size between bacteria

Genomes of Bacteria – Large Variation

- *Tremblaya princeps* 140kb, 121 coding sequences
- *Sorangium cellulosum* 14000kb , 11599 coding sequences

Genomes of Bacteria;

Comparison of regulatory genes in bacterial genomes			
Microorganism	# Genes in the Genome	# Regulatory Proteins	% of Total
<i>Pseudomonas aeruginosa</i>	5570	468	8.4
<i>Escherichia coli</i>	4289	250	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

Distribution of genes among selected bacterial genomes and their sizes;

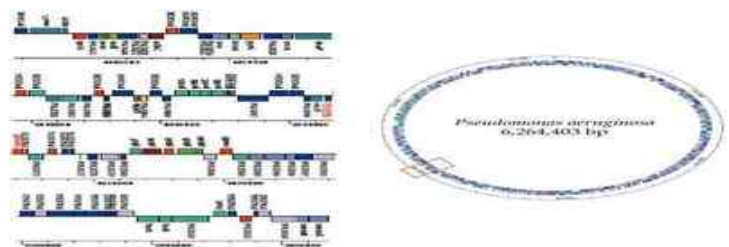
Organism	Genome Size (Mbp)	No. of ORFs (% coding)	Unknown Function	Unique ORFs
<i>Aeropyrum pernix</i> K1	1.67	1,885 (89%)		
<i>A. aeolicus</i> VF5	1.50	1,749 (93%)	663 (44%)	407 (27%)
<i>A. fulgidus</i>	2.18	2,437 (92%)	1,315 (54%)	641 (26%)
<i>B. subtilis</i>	4.20	4,779 (87%)	1,722 (42%)	1,053 (26%)
<i>B. burgdorferi</i>	1.44	1,738 (88%)	1,132 (65%)	682 (39%)
<i>Chlamydia pneumoniae</i> AR39	1.23	1,134 (90%)	543 (48%)	262 (23%)
<i>Chlamydia trachomatis</i> MoP ₈	1.07	936 (91%)	353 (38%)	77 (8%)
<i>C. trachomatis</i> serovar D	1.04	928 (92%)	290 (32%)	255 (29%)
<i>Deinococcus radiodurans</i>	3.28	3,187 (91%)	1,715 (54%)	1,001 (31%)
<i>E. coli</i> K-12-MG1655	4.60	5,295 (88%)	1,632 (38%)	1,114 (26%)
<i>H. influenzae</i>	1.83	1,738 (88%)	595 (35%)	237 (14%)
<i>H. pylori</i> 26695	1.66	1,589 (91%)	744 (45%)	539 (33%)
<i>Methanobacterium thermoautotrophicum</i>	1.75	2,008 (90%)	1,010 (54%)	496 (27%)

Organism	Genome Size (Mbp)	No. of ORFs (% coding)	Unknown Function	Unique ORFs
<i>Methanococcus jannaschii</i>	1.66	1,783 (87%)	1,076 (62%)	525 (30%)
<i>M. tuberculosis</i> H37Rv	4.41	4,275 (92%)	1,521 (39%)	606 (15%)
<i>M. genitalium</i>	0.58	483 (91%)	173 (37%)	7 (2%)
<i>M. pneumoniae</i>	0.81	680 (89%)	248 (37%)	67 (10%)
<i>N. meningitidis</i> MC58	2.24	2,155 (83%)	856 (40%)	517 (24%)
<i>Pyrococcus horikoshii</i> OT3	1.74	1,994 (91%)	589 (42%)	453 (22%)
<i>Rickettsia prowazekii</i> Madrid E	1.11	878 (75%)	311 (37%)	209 (25%)
<i>Sporobolus</i> sp.	3.57	4,003 (87%)	2,384 (75%)	1,426 (45%)
<i>T. maritima</i> MSB8	1.86	1,879 (95%)	863 (46%)	373 (26%)
<i>T. pallidum</i>	1.14	1,039 (93%)	461 (44%)	280 (27%)
<i>Vibrio cholerae</i> El Tor N1696	4.03	3,890 (88%)	1,806 (46%)	934 (24%)
	50.60	52,462 (89%)	22,358 (43%)	12,161 (23%)

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

Yeast Genome - Characteristics

- Small and compact
- Small intergenic sequences

(DNA sequence between two genes is called as intergenic DNA sequence.)

- Few transposable elements
- Few introns
- Limited RNA interference
- The yeast genome is predicted to contain about 6,200 genes
- 274 tRNA
- 287 introns
- Small percentage of yeast genes have introns
- The intergenic space between genes is only between 200bp - 1,000bp

Yeast Genome: Genome of Yeast Cell;

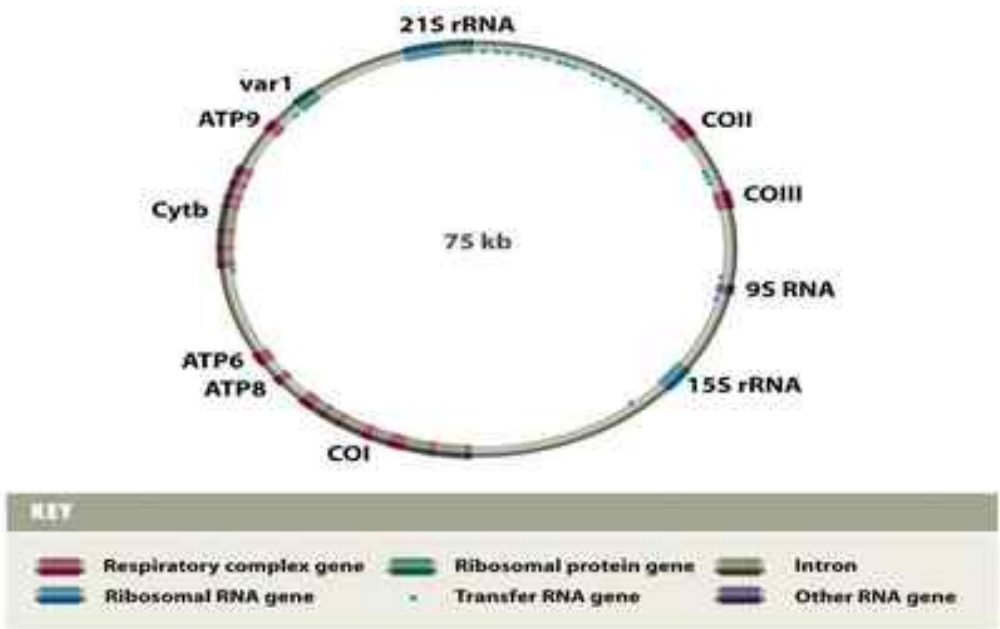
Characteristic	Chromosomes	Plasmid	Mitochondria
Relative amount (%)	85	5	10
Number of copies	2 x 16	60-100	~50 (8-130)
Size (kbp)	~ 12,100	6.318	70-76

- The largest known regulatory sequences are spread over about 2,800bp e.g. MUC1/FLO11 (name of the gene)
- Yeast genes have names consisting of three letters and up to three numbers e.g. GPD1, HSP12, PDC6
- Usually they are meaningful

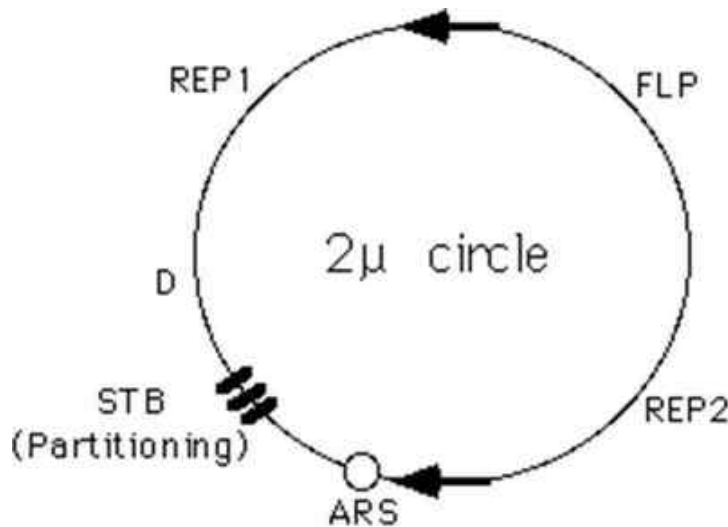
Yeast Genome – Genes Nomenclature

- Wild type genes are written with capital letters in italics: *TPS1*, *RHO1*, *CDC28*
- Recessive mutant genes are written with small letters in italics: *tps1*, *rho1*, *cdc28*
- Three letters provides information about a function, mutant phenotype, or process related to that gene.
- CDC - Cell Division Cycle ; ADE-ADenine biosynthesis

Yeast – Mitochondrial DNA;



- 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 Genome – Conclusion

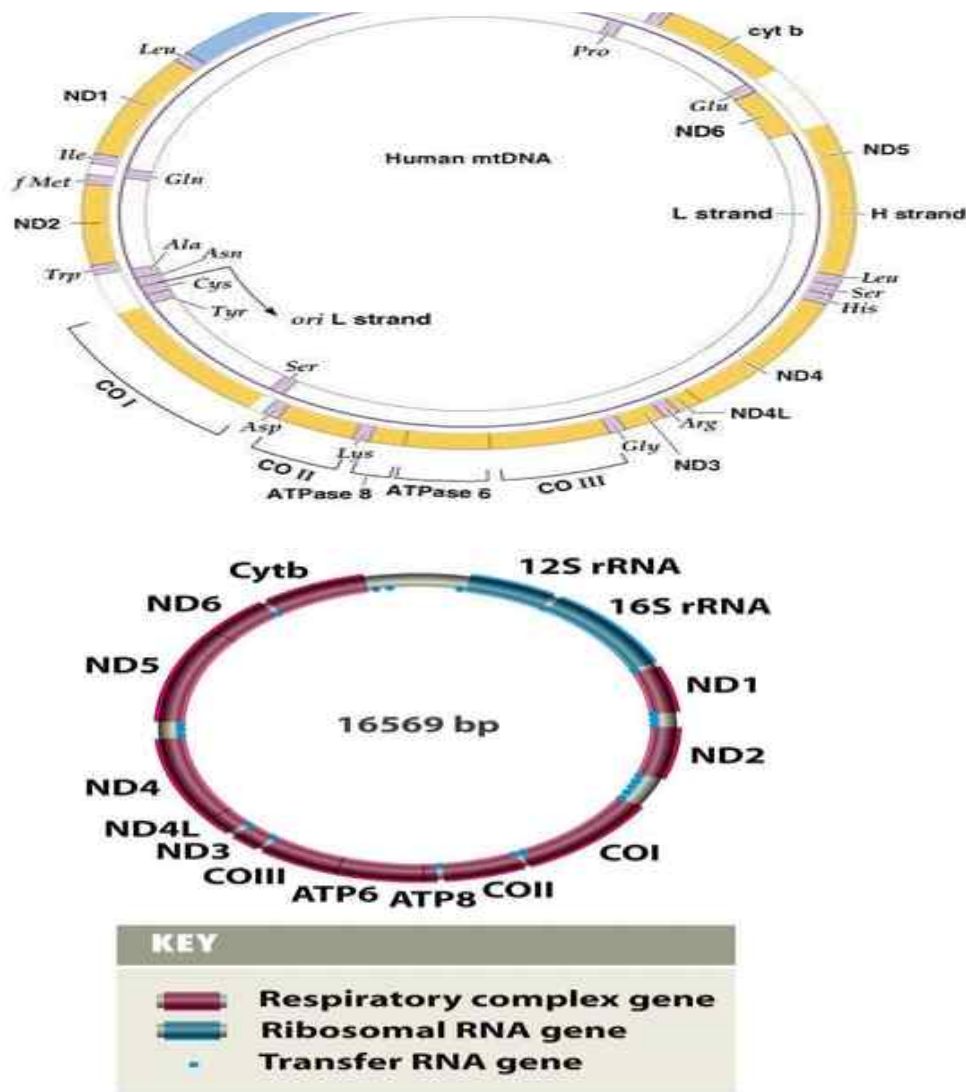
- Yeast nuclear genome has 16 chromosomes.
- A mitochondrial genome.
- A plasmid

42.Mitochondrial Genome:

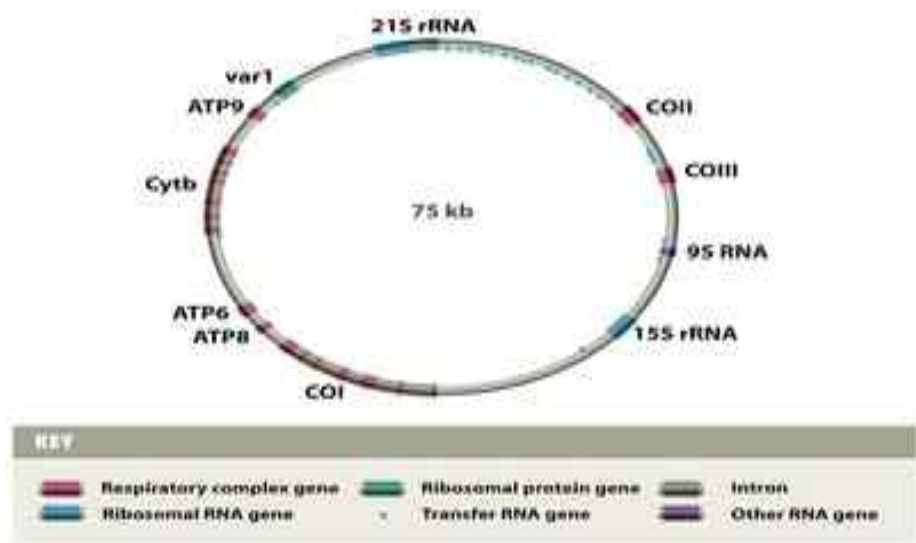
- Multiple identical circular chromosomes
- ~15-16 Kb in animals
- ~ 200 kb to 2,500 kb in plants
- Over 95% of mitochondrial proteins are encoded in the nuclear genome.
- Often A+T rich genomes

Human Mitochondrial Genome

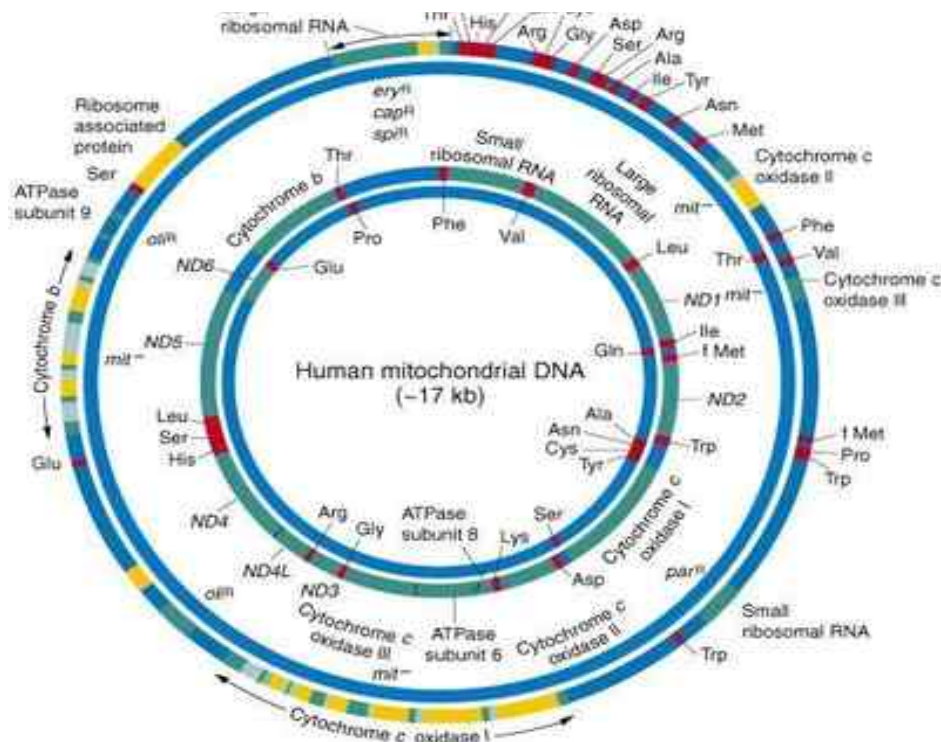
- Circular, double stranded, 16.6 kb
- The two strands are notably different in base composition, leading to one strand being heavy (H strand) and the other light (L strand).
- Both strands encode genes, although more are on the H strand.
- A short region (1121 bp), the D loop is a DNA triple helix: 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)
- Human mitochondrial chromosome consists of 37 Genes. Out of which 22 are tRNAs, 2 are rRNAs and 13 are polypeptides.
- tRNA: only 60 of the 64 codons code for amino acids.



Mitochondrial Genome – Yeast;



Mitochondrial Genome – Human and Yeast;



Mitochondrial Genomes - Comparison

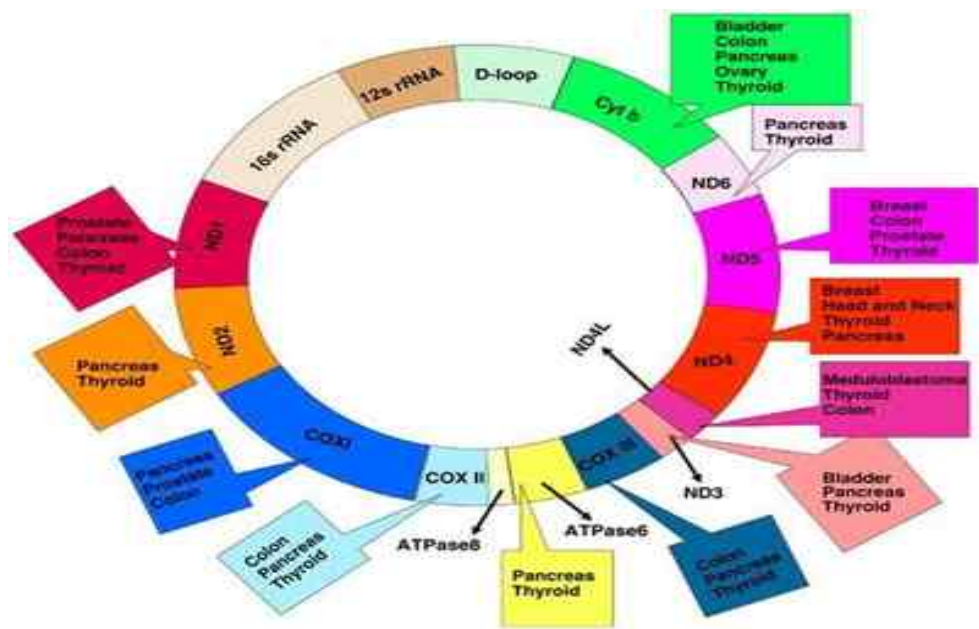
Species	Type of organism	Genome size (kb)
Mitochondrial genomes		
<i>Plasmodium falciparum</i>	Protozoan (malaria parasite)	6
<i>Chlamydomonas reinhardtii</i>	Green alga	16
<i>Mus musculus</i>	Vertebrate (mouse)	16
<i>Homo sapiens</i>	Vertebrate (human)	17
<i>Metridium senile</i>	Invertebrate (sea anemone)	17
<i>Drosophila melanogaster</i>	Invertebrate (fruit fly)	19
<i>Chondrus crispus</i>	Red alga	26
<i>Aspergillus nidulans</i>	Ascomycete fungus	33
<i>Reclinomonas americana</i>	Protozoa	69
<i>Saccharomyces cerevisiae</i>	Yeast	75
<i>Suillus griseellus</i>	Basidiomycete fungus	121
<i>Brassica oleracea</i>	Flowering plant (cabbage)	160
<i>Arabidopsis thaliana</i>	Flowering plant (vetch)	367
<i>Zea mays</i>	Flowering plant (maize)	570
<i>Cucumis melo</i>	Flowering plant (melon)	2500

Feature	<i>Plasmodium falciparum</i>	<i>Chlamydomonas reinhardtii</i>	<i>Homo sapiens</i>	<i>Saccharomyces cerevisiae</i>	<i>Arabidopsis thaliana</i>	<i>Reclinomonas americana</i>
Total number of genes	5	12	37	35	52	92
Types of genes						
Protein-coding genes	3	7	13	8	27	62
Respiratory complex	3	7	13	7	17	24
Ribosomal proteins	0	0	0	1	7	27
Transport proteins	0	0	0	0	3	6
RNA polymerase	0	0	0	0	0	4
Translation factor	0	0	0	0	0	1
Functional RNA genes	2	5	24	27	25	30
Ribosomal RNA genes	2	2	2	2	3	3
Transfer RNA genes	0	3	22	24	22	26
Other RNA genes	0	0	0	1	0	1
Number of introns	0	1	0	8	23	1
Genome size (kb)	6	16	17	75	367	69

		Second letter				
		U	C	A	G	
First letter	U	Phe	Ser	Tyr	Cys	U
	Phe	Ser	Tyr	Cys	Cys	C
	Leu	Ser	Stop	Stop	Stop	A
	Leu	Ser	Stop	Stop	Stop	G
C	Leu	Pro	His	Arg	Arg	U
	Leu	Pro	His	Arg	Arg	C
	Leu	Pro	Gln	Arg	Arg	A
	Leu	Pro	Gln	Arg	Arg	G
A	Se (Met)	Thr	Asn	Ser	Ser	U
	Se (Met)	Thr	Asn	Ser	Ser	C
	Se (Met)	Thr	Lys	(Arg) Stop	(Arg) Stop	A
	Se (Met)	Thr	Lys	(Arg) Stop	(Arg) Stop	G
G	Val	Ala	Asp	Gly	Gly	U
	Val	Ala	Asp	Gly	Gly	C
	Val	Ala	Glu	Gly	Gly	A
	Val	Ala	Glu	Gly	Gly	G

		Second letter				
		U	C	A	G	
First letter	U	UUU Phe	UUC Ser	UAU Tyr	UGU Cys	U
	Phe	UUC Ser	UUA Leu	UAA Stop	UGC Cys	C
	Leu	UUA Leu	UUG Leu	UAG Stop	UGA Stop	A
	Leu	UUG Leu			UGG Trp	G
C	Leu	CCU Pro	CCA Pro	CAU His	CGU Arg	U
	Leu	CCC Pro	CCG Pro	CAC His	CGC Arg	C
	Leu	CCA Pro	CCG Pro	CAA Gln	CGA Arg	A
	Leu	CCG Pro		CAG Gln	CGG Arg	G
A	Asn	AAU Asn	AAU Asn	AAU Asn	AAU Asn	U
	Asn	AAC Asn	AAC Thr	AAC Asn	AAC Asn	C
	Lys	AAA Lys	ACA Thr	AAA Lys	AGA Arg	A
	Lys	AAG Lys	ACG Thr	AAG Lys	AGG Arg	G
G	Asp	GAU Asp	GCU Ala	GAU Asp	GAU Asp	U
	Asp	GUC Asp	GCC Ala	GAC Asp	GAC Asp	C
	Gly	GUA Gly	GCA Gly	GAA Gly	GGA Gly	A
	Gly	GUG Gly	GCG Gly	GAG Gly	GGG Gly	G

Mitochondrial Genome – Tumors due to mutations



Mitochondrial DNA

polymorphisms track human migrations;

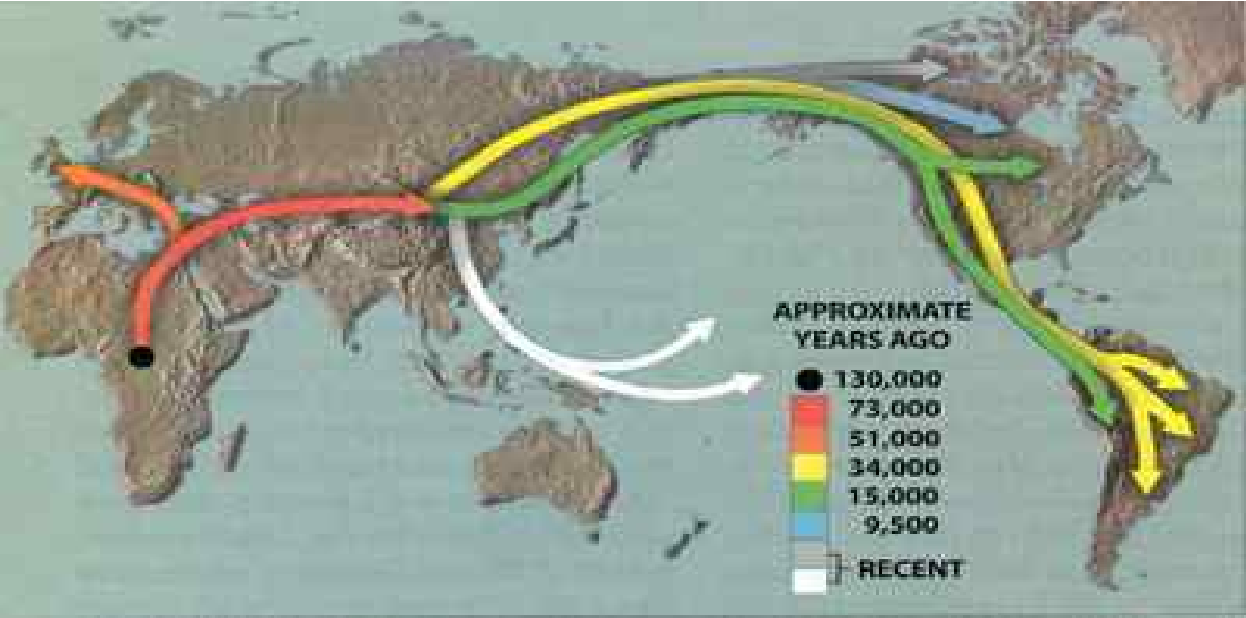


Fig: All humans descend from a small group of Africans

This group originated in central Africa ~200,000 years ago

2 4

The founding group was small (10 -10 people)

Conclusion

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

- 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).

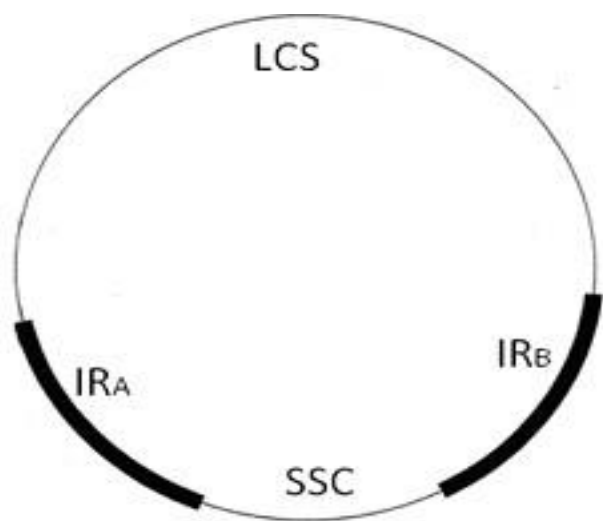
Chloroplast Genomes: Size in different Taxa;

Taxa	Genome size (In kb)	Inverse duplication (In kb)
Angiospermae		
<i>Nicotiana tabacum</i>	156	25
<i>Spinacia oleracea</i>	150	24
<i>Pelargonium hortorum</i>	217	76
<i>Pisum sativum</i>	120	Not present
<i>Epifragus virginiana</i>	70	22
<i>Oryza sativa</i>	134	21
Gymnospermae		
<i>Pinus</i>	120	Not present
<i>Ginkgo biloba</i>	158	17
Pteridophyta		
<i>Osmunda cinnamomea</i>	144	10
Bryophyta		
<i>Marchantia polymorpha</i>	121	10
Chlorophyta		
<i>Codium fragile</i>	85	Not present
<i>Chlamydomonas reinhardtii</i>	195	22
<i>Chlamydomonas moewusii</i>	292	41
Rhodophyta		
<i>Cyanophora paradoxa</i>	127	10
Chromophyta		
<i>Dictyota dichotoma</i>	123	5

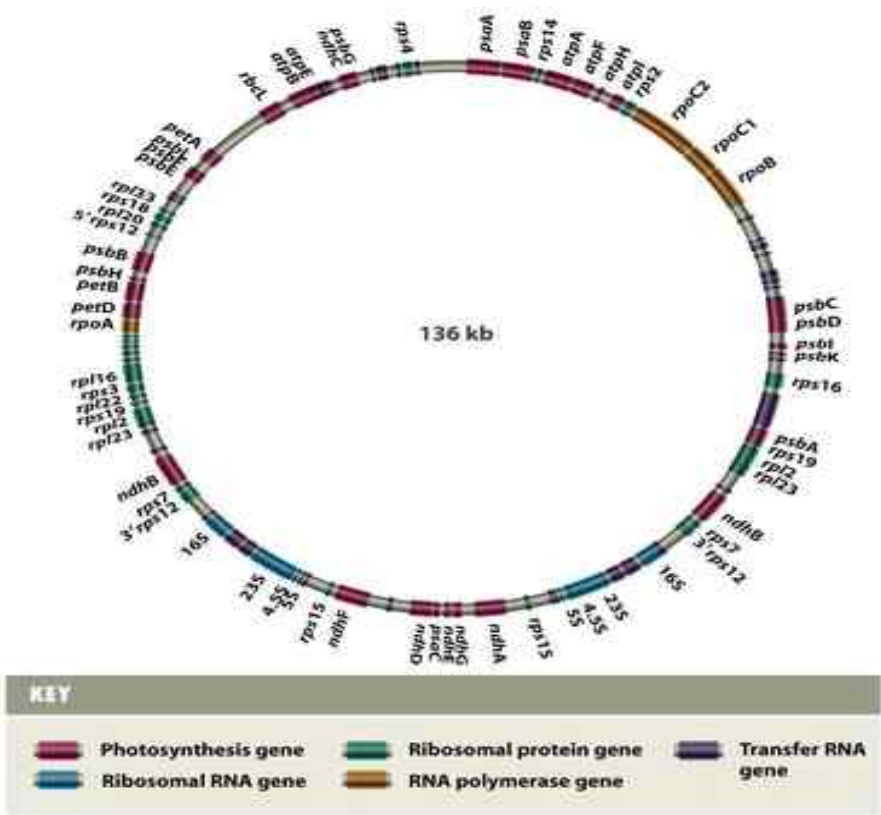
Chloroplast Genome Size;

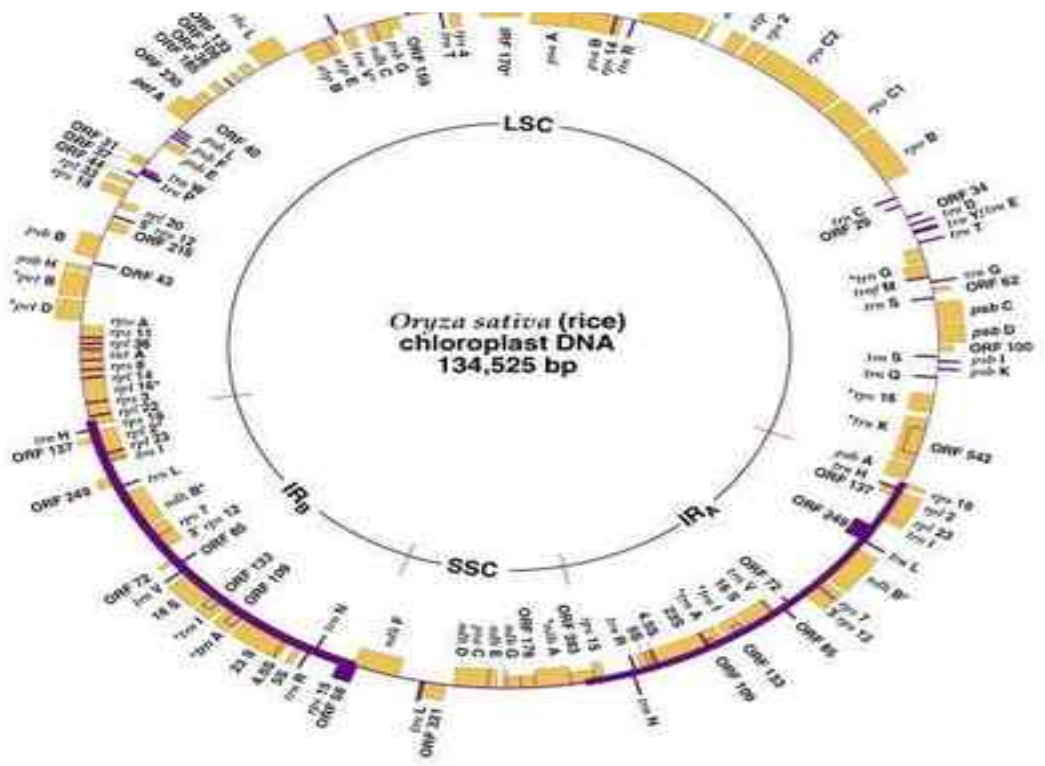
Species	Type of organism	Genome size (kb)
Chloroplast genomes		
<i>Pisum sativum</i>	Flowering plant (pea)	120
<i>Marchantia polymorpha</i>	Liverwort	121
<i>Oryza sativa</i>	Flowering plant (rice)	136
<i>Nicotiana tabacum</i>	Flowering plant (tobacco)	156
<i>Chlamydomonas reinhardtii</i>	Green alga	195

- Large Single Copy Region
- Small Single Copy Region
- Inverted Repeats



Chloroplast Genome of Rice;





Functions of Chloroplast Genes

- Most cp genes fall into two functional groups:
- (i) Genes involved in replication, transcription, translation
- (ii) 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.)

Properties of Chloroplast Genome

- 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.

Chloroplast Genome - Conclusion

- 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 (Organelles genomes can be in the form of ‘Mitochondrial genome’ or ‘Chloroplast genome’.
- Contains repetitive sequences
- SINEs (short interspersed elements)
- LINEs (long interspersed elements)

LINES and SINES

- **LINES**
- 1-5 kb

- 200-300 bp
- 100,000 copies

Highly repetitive DNA sequences present in eukaryotic genomes can be of different types. i.e.

- Minisatellites
- Microsatellites
- Telomeres

Minisatellites

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

Microsatellites

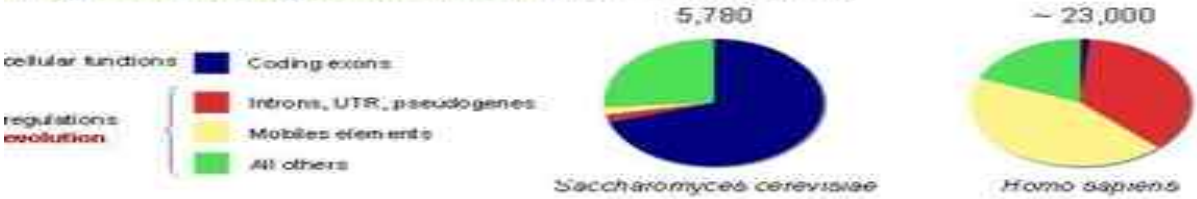
- Repeats up to 13 bp

Telomeres

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

Elements of Eukaryotic Genomes;

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



Comparison - Prokaryotic & Eukaryotic Genomes

Prokaryotic	Eukaryotic
• Usually circular • Smaller • Found in the nucleoid region • Less elaborately structured and folded	• 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

45.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 vary in genes numbers;

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

- Nematode
C. elegans has 100 Mb and 20,000 genes, while *Drosophila* has 165 Mb and 13,700 genes
- Vertebrate genomes can produce more than one polypeptide per gene because of alternative splicing of RNA transcripts

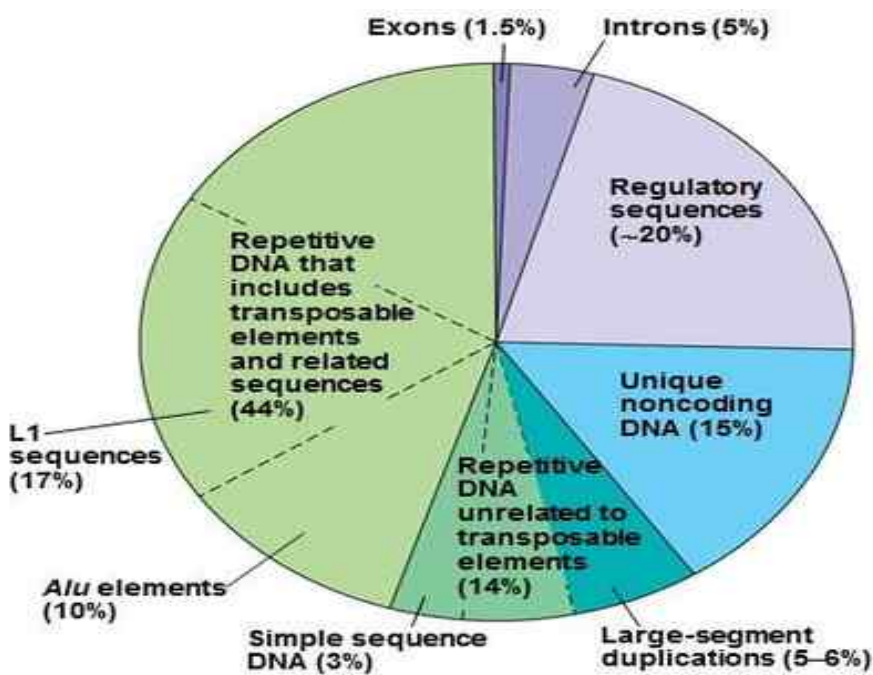
Humans and Mammals have low gene density;

- 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

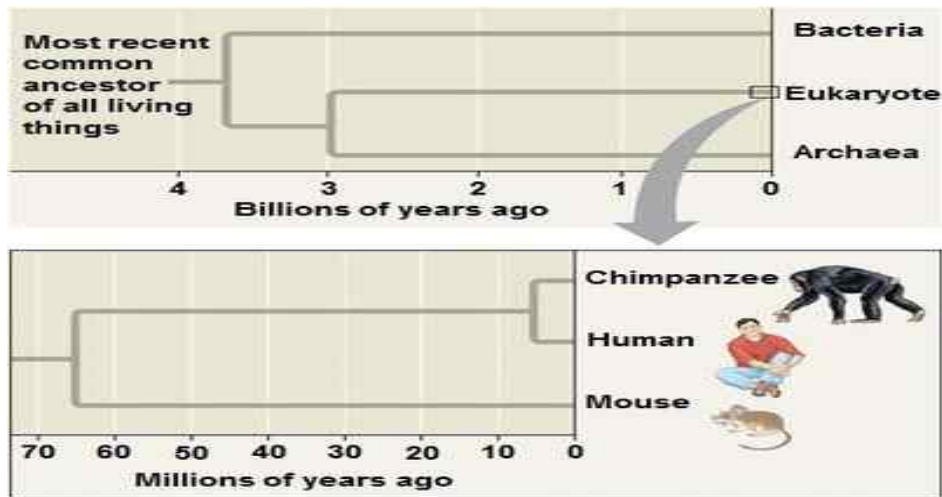
Human Genome: Distribution of coding and non-coding DNA;



Comparing Genomes

- 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 (Figure)



Genome size	Most are 1–6 Mb		most are 10–100 Mb, but a few are much larger
Number of genes	1,500–7,500		5,000–40,000
Gene density	Higher than in eukaryotes		Lower than in prokaryotes (Within eukaryotes, lower density is correlated with larger genomes.)
Introns	None in protein-coding genes	Present in some genes	Unicellular eukaryotes: present, but prevalent only in some species Multicellular eukaryotes: present in most genes
Other noncoding DNA	Very little		Can be large amounts; generally more repetitive noncoding DNA in multicellular eukaryotes

Comparing Genomes;

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

46.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.

Comparing closely related species:

- 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 differs from each other.
- Eukaryotes and prokaryotes genomes differ very significantly.
- Size of genomes - 1000 folds difference between eukaryotes and prokaryotes.
- ~ 30 fold between genomes of different eukaryotes.
- In humans ~ 23,000
- Bacterial genomes ~ 1,500 – 2,000 genes.

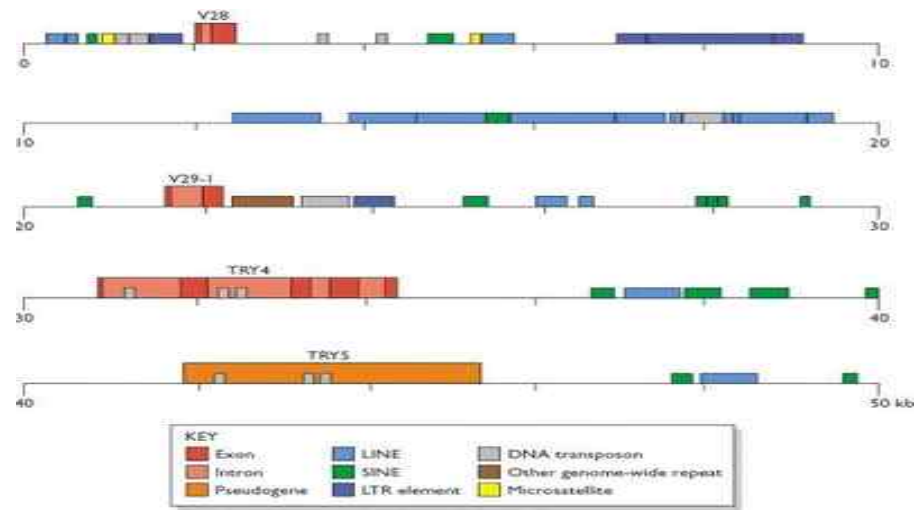
Genome Anatomy - Eukaryotes

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

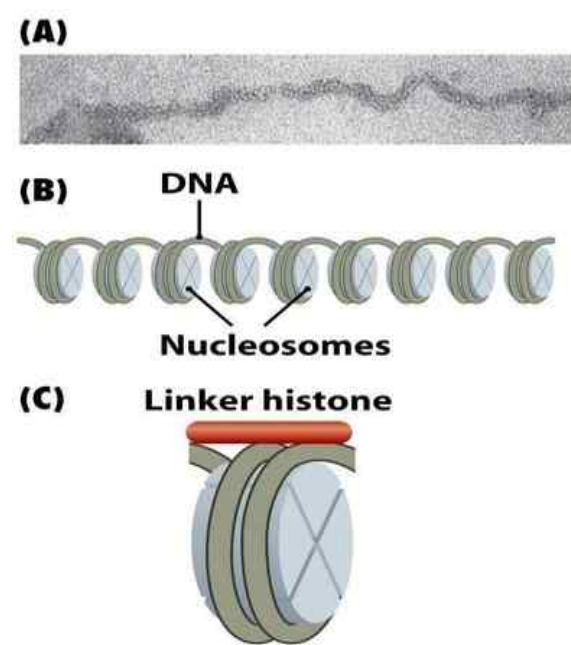
Genome Anatomy - Prokaryotes

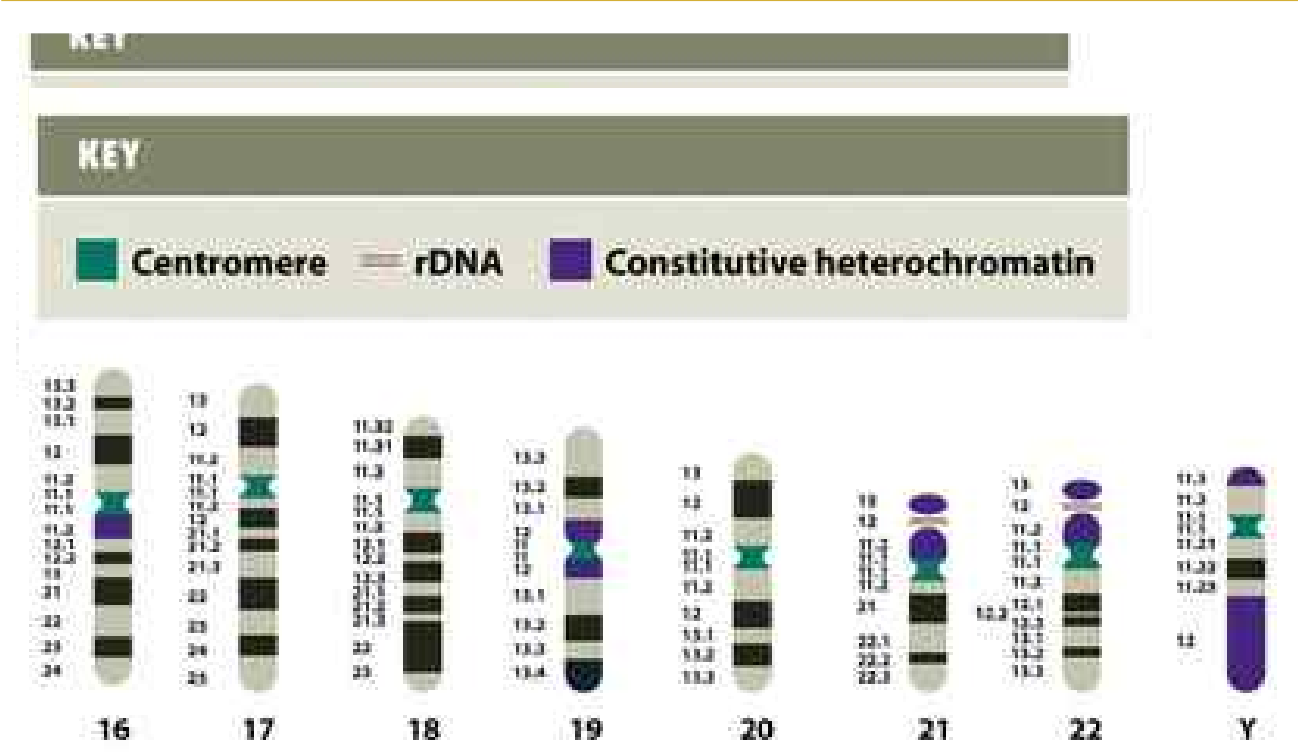
- Prokaryotes have a few repeats and transposable elements and their genomes consist of mainly the genes.

2 Gene Segments: V28 & V29-1
1 Pseudogene: TRY5
52 genome wide repeat sequences:
LINE, SINE, LTR, & DNA transposon.
Two Microsatellites



Genome Organization;





Species	DNA molecules	Genome organization	
		Size (Mb)	Number of genes
<i>Escherichia coli</i> K12	One circular molecule	4.639	4405
<i>Vibrio cholerae</i> El Tor N16961	Two circular molecules		
	Main chromosome	2.961	2770
	Megaplasmid	1.075	1115
<i>Deinococcus radiodurans</i> R1	Four circular molecules		
	Chromosome 1	2.649	2653
	Chromosome 2	0.412	369
	Megaplasmid	0.177	145
	Plasmid	0.046	40

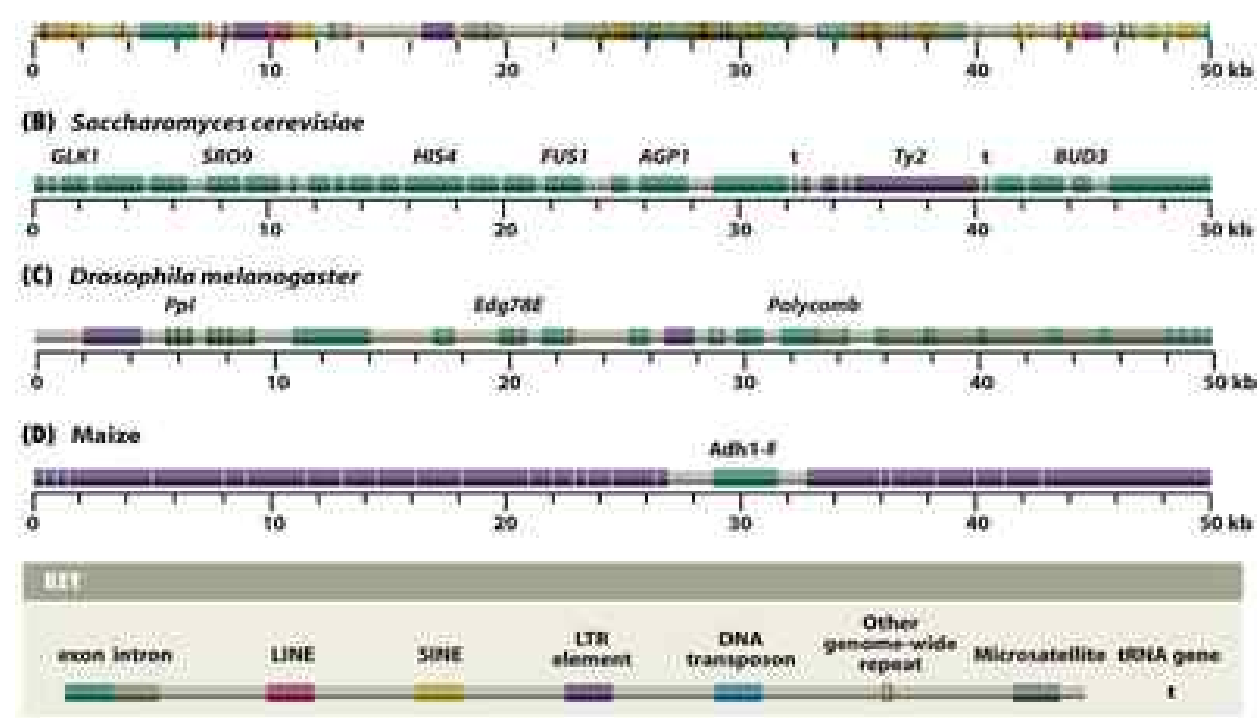
Species	Size of genome (Mb)	Approximate number of genes
Bacteria		
<i>Mycoplasma genitalium</i>	0.58	500
<i>Streptococcus pneumoniae</i>	2.16	2300
<i>Vibrio cholerae</i> El Tor N16961	4.03	4000
<i>Mycobacterium tuberculosis</i> H37Rv	4.41	4000
<i>Escherichia coli</i> K12	4.64	4400
<i>Yersinia pestis</i> CO92	4.65	4100
<i>Pseudomonas aeruginosa</i> PAO1	6.26	5700
Archaea		
<i>Methanococcus jannaschii</i>	1.66	1750
<i>Archaeoglobus fulgidus</i>	2.18	2500

Genome

Organization in Prokaryotes;
Genome Organization:
Comparisons

Species	Genome size (Mb)
Vertebrates	
<i>Takifugu rubripes</i> (pufferfish)	400
<i>Homo sapiens</i>	3200
<i>Mus musculus</i> (mouse)	3300
Plants	
<i>Arabidopsis thaliana</i> (vetch)	125
<i>Oryza sativa</i> (rice)	466
<i>Zea mays</i> (maize)	2500
<i>Pisum sativum</i> (pea)	4800
<i>Triticum aestivum</i> (wheat)	16,000
<i>Fritillaria assyniaca</i> (fritillary)	120,000

Genome



Organization: Human, Yeast, Fruit Fly , Maize (Comparison)

Compactness of Genomes;

Feature	Yeast	Fruit Fly	Human
Gene density (average number per Mb)	496	26	11
Introns per gene (average)	0.04	3	9
Amount of the genome that is taken up by genome-wide repeats	3.4%	12%	44%

Species	Genome size (Mb)
Fungi	
<i>Saccharomyces cerevisiae</i>	12.1
<i>Aspergillus nidulans</i>	25.4
Protozoa	
<i>Tetrahymena pyriformis</i>	190
Invertebrates	
<i>Caenorhabditis elegans</i>	97
<i>Drosophila melanogaster</i>	180
<i>Bombyx mori</i> (silkworm)	490
<i>Strongylocentrotus purpuratus</i> (sea urchin)	845
<i>Locusta migratoria</i> (locust)	5000

Conclusion

- Anatomy of different genomes differs 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:

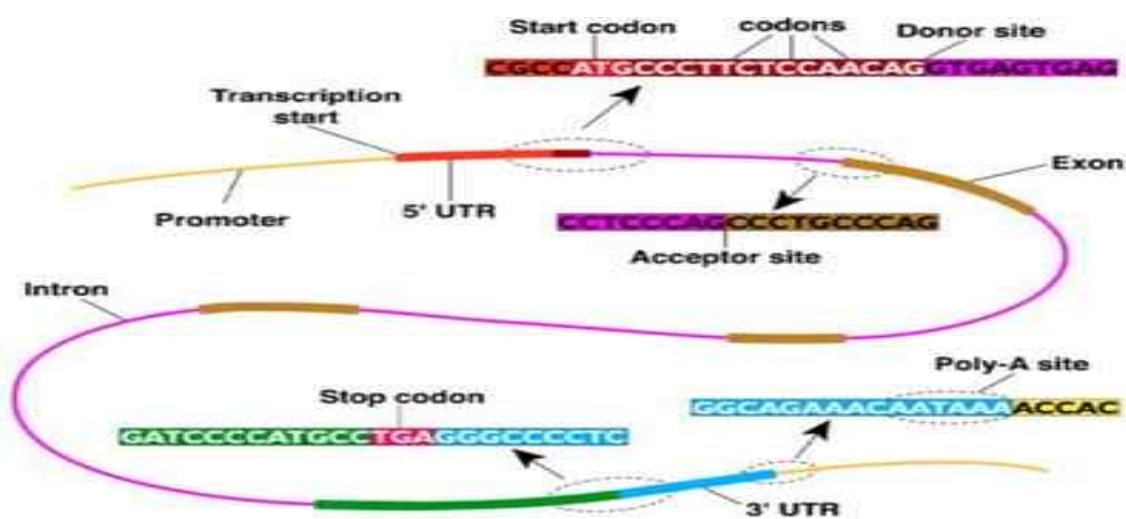
- Segments of DNA that encode proteins or other functional products are called genes.
- Gene sequences are transcribed into messenger RNA (mRNA).

- A gene consists of three main components;
- (i) Open Reading Frame: From start codon (ATG) to stop (TGA, TAA, TAG) • (ii) Upstream region with binding site. (e.g. TATA box, GC box, CAAT box). • (iii) Poly-a tail (Present at the 3 end of the gene).

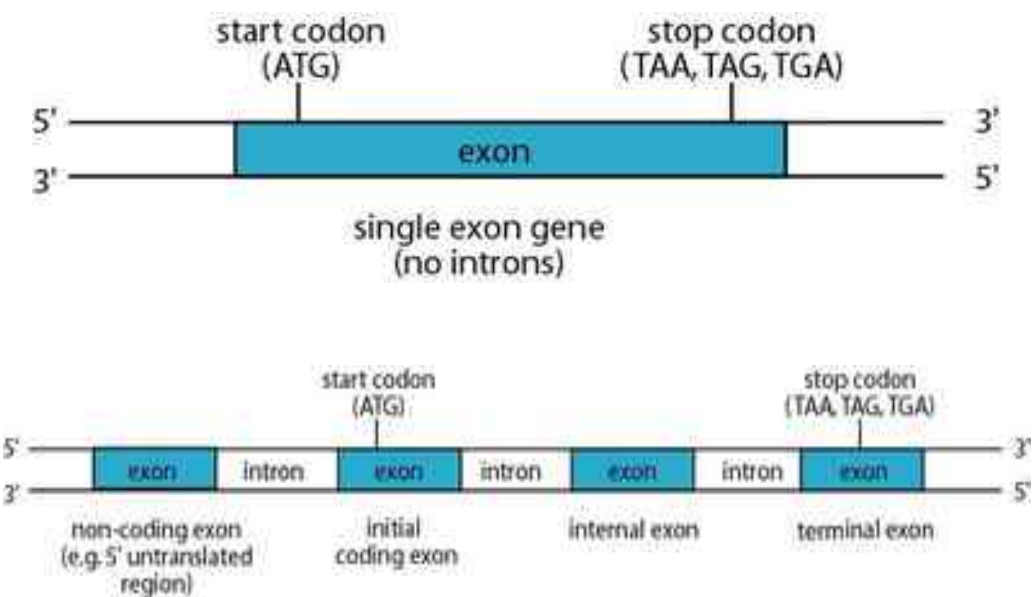
Gene Anatomy – Typical Prokaryotic Gene



Gene Anatomy – Typical Eukaryotic Gene



Single Exon Gene and Multiple Exons Gene (Comparison)



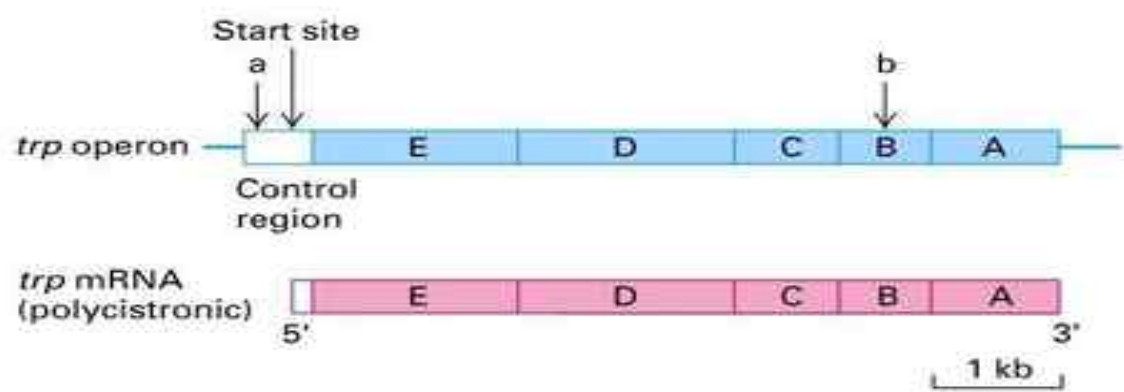
What is Gene – Conclusion

- 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:

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
- Polycistronic mRNA encodes several proteins

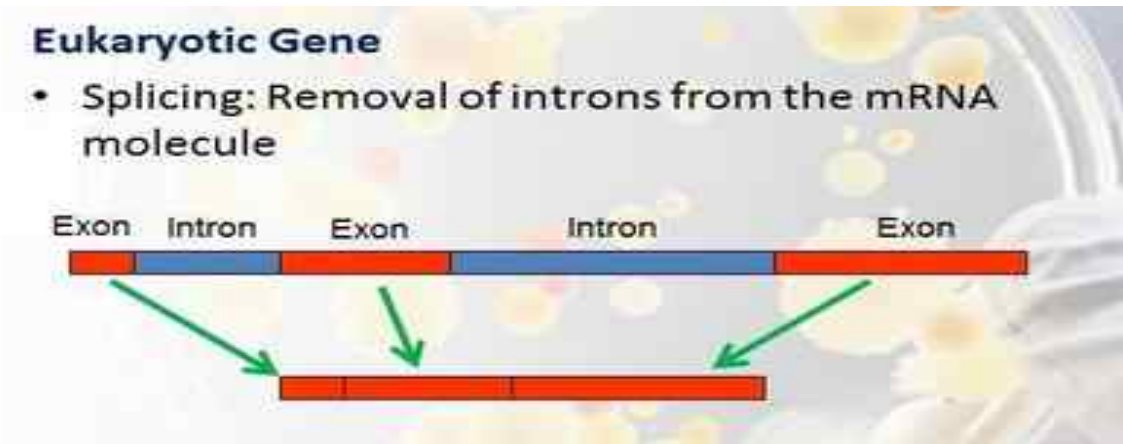


Eukaryotic Gene: Exons and Introns

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

In Eukaryotic genes we also see a phenomenon which is called splicing;

- Splicing: Removal of introns from the mRNA molecule.
- Once mRNA is produced from the DNA then intronic sequences are removed from the sequence of mRNA. This phenomenon is called as splicing.

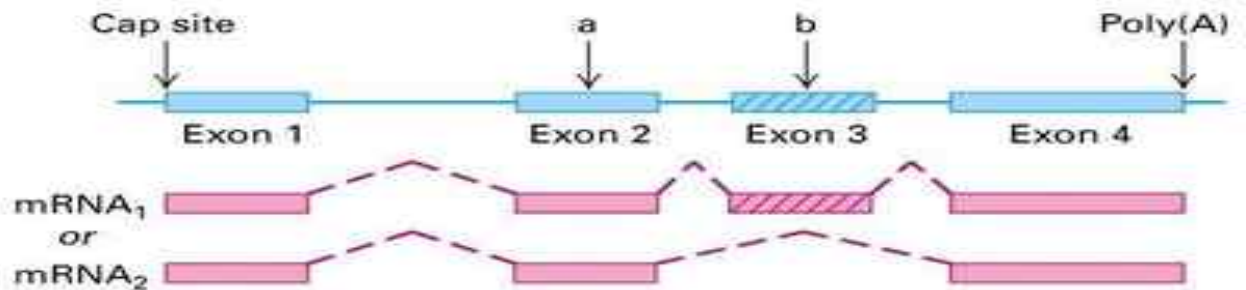


Eukaryotic Gene usually have;

- Organize expression of genes' (function calls)
- Promoter region (binding site), usually near coding region
- Binding can block/inhibit expression

There are sequences near to binding site which sometime enhance the expression. Those molecules are called as enhancers. If they stop the expression then they are called silencers.

Alternative Splicing in Eukaryotic Genes



- Large in size

Eukaryotic Gene: Alternative Splicing (Figure)

Eukaryotic Gene is complex in nature and to identify or to work on it is difficult. Computational challenges identify binding sites correlate sequence to expression.

Conclusion

- Most of prokaryotic genes are without introns and they 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
- Long interspersed elements
- Short interspersed elements

- 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 elements)
 - Transposons
 - Viral retrotransposons
 - Long interspersed elements (LINES; nonviral retrotransposons)
 - Short interspersed elements (SINES; nonviral retrotransposons)
- Unclassified spacer DNA

Major Classes of Eukaryotic DNA in Human Genome;

Class	Length	Copy Number in Human Genome	Fraction of Human Genome, %
Protein-coding genes			
Solitary genes	Variable	1	$\approx 1.5^{\circ} (0.8)^{\circ}$
Duplicated or diverged genes in gene families	Variable	2–1000	$\approx 1.5^{\circ} (0.8)^{\circ}$
Tandemly repeated genes encoding rRNAs, tRNAs, mRNAs, and histones	Variable	20–300	0.3
Repetitious DNA			
Simple-sequence DNA	1–600 bp	Variable	1
Interspersed repeats			
DNA transposons	2–3 kb	300,000	1
LTR retrotransposons	6–11 kb	440,000	8
Non-LTR retrotransposons			
LINEs	6–8 kb	860,000	21
SINEs	100–300 bp	1,600,000	13
Processed pseudogenes	Variable	1–100	≈ 0.4
Unclassified spacer DNA	Variable	n.a. ¹	$\approx 2\%$

Types of Eukaryotic DNA – Conclusion

- Protein coding genes
- Tandemly repeated genes
- Repeated DNA
- Unclassified spacer DNA

51.Duplicated Genes and Pseudo-Genes:

Duplicated Genes

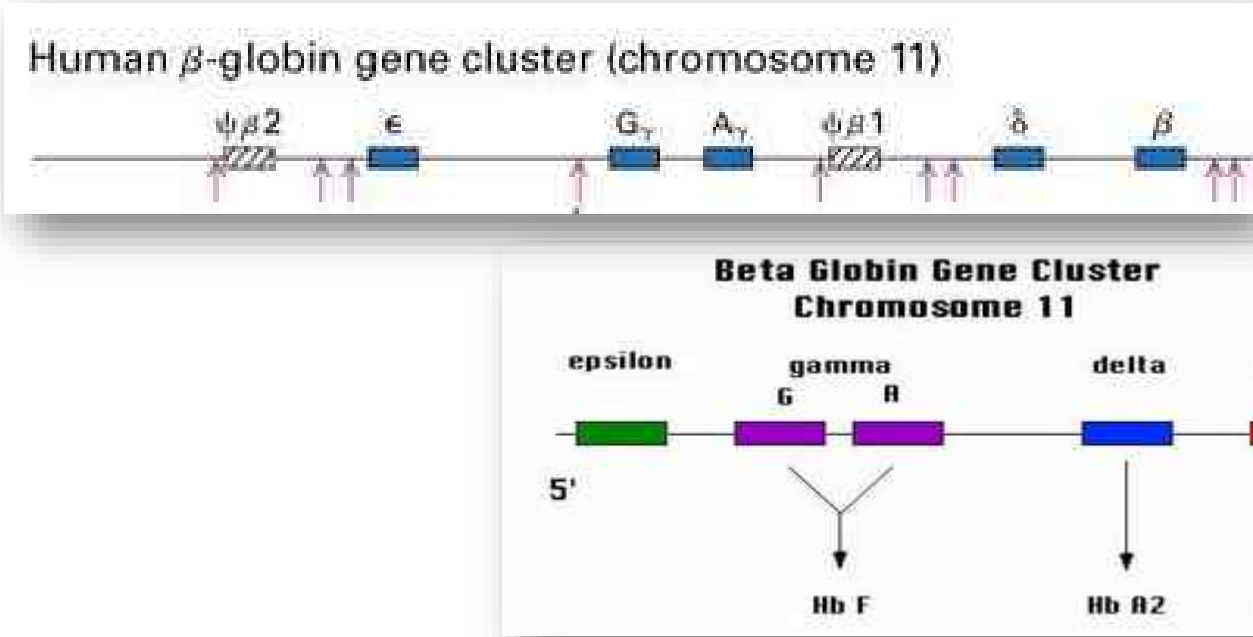
- Encode closely related (homologous) proteins
- Clustered together in genome
- Formed by duplication of an ancestral gene followed by mutation

Pseudo-Genes

- Nonfunctional copies of genes • Formed by duplication of ancestral gene, or reverse transcription and integration.

Duplicated Genes (Figure)

Five functional genes and two pseudo-genes



Pseudo-Genes

- Not expressed due to mutations that produce;
- A stop codon
- Prevent mRNA processing
- Due to lack of regulatory sequences

Pseudo-Genes (Figure)

- Five functional genes and two pseudo-genes



Gene's duplication: Globin family

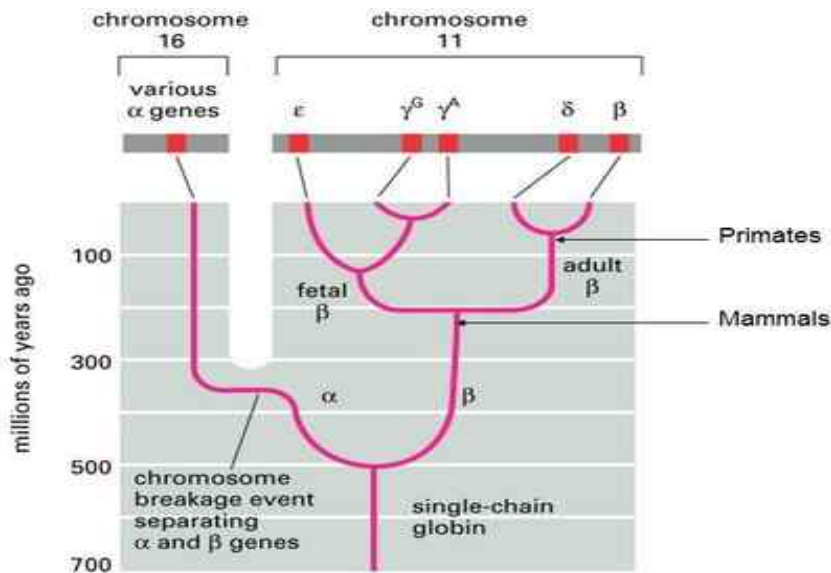
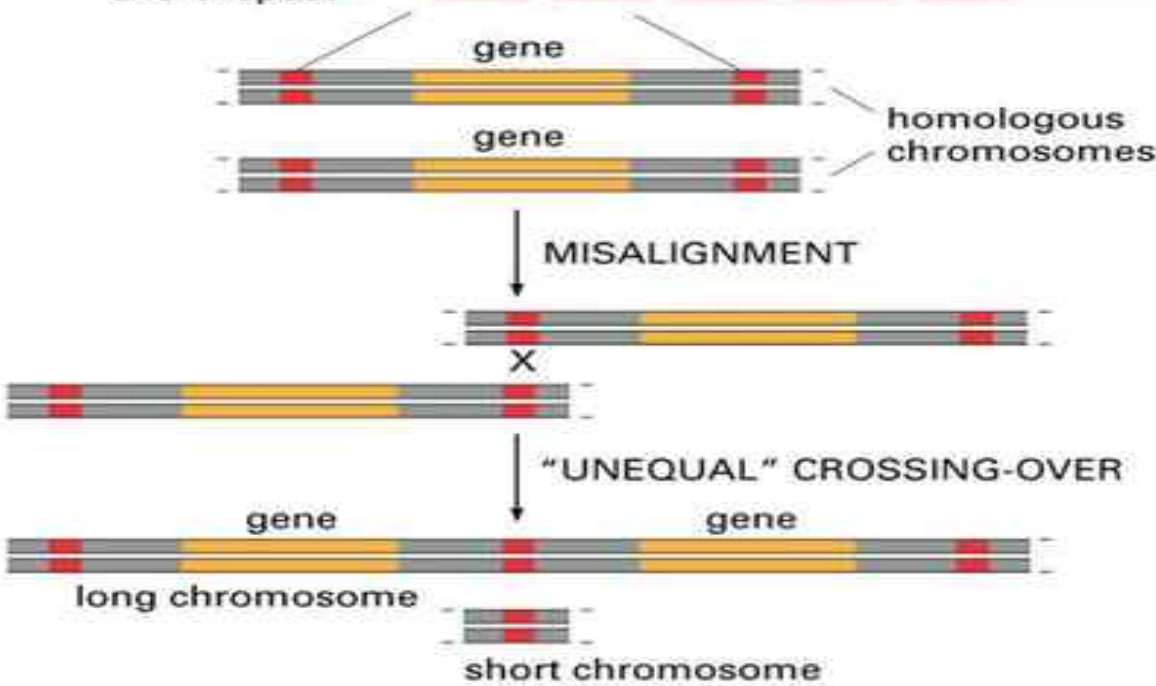
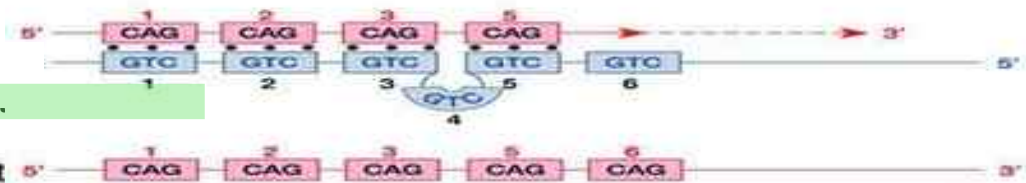
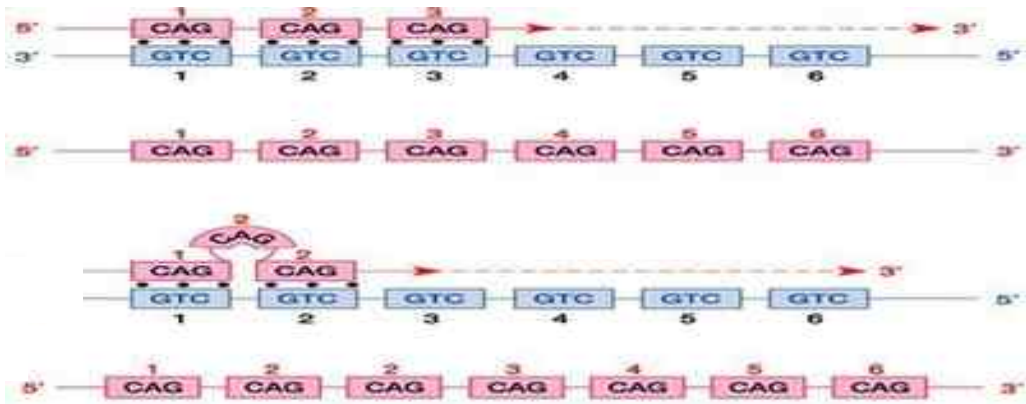
In case of human the example of genes duplication is Globin family.

- 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

Normal replication

Backward slippage causes insertions

Forward slippage causes deletions

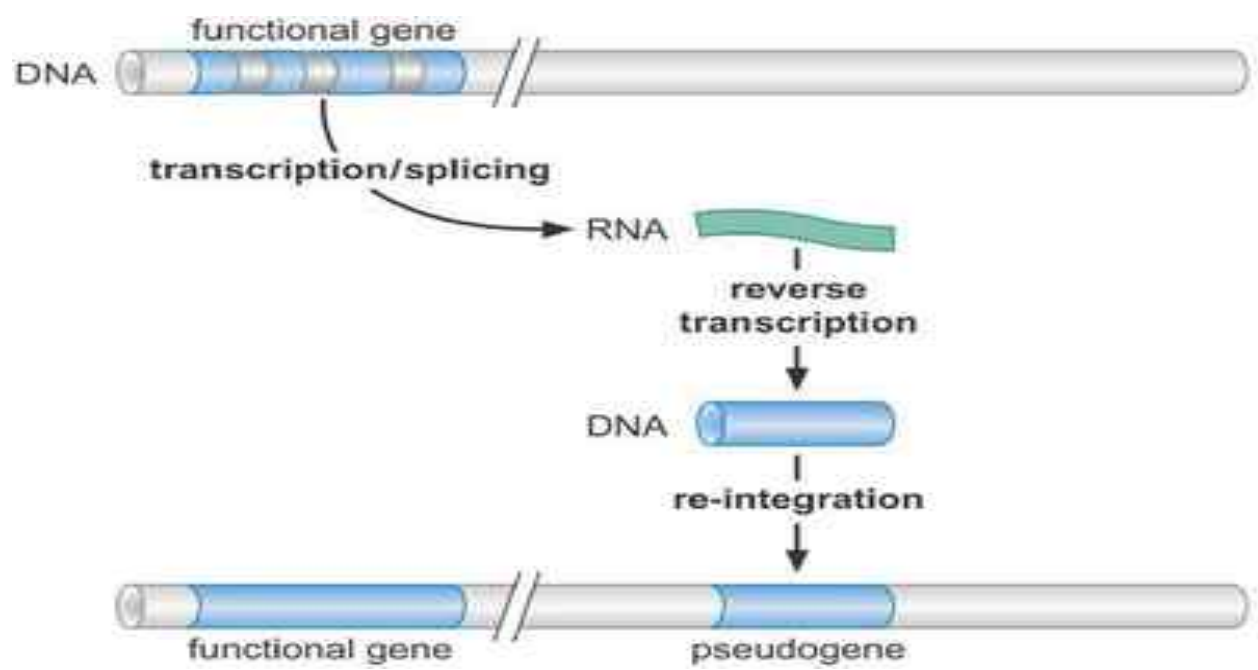


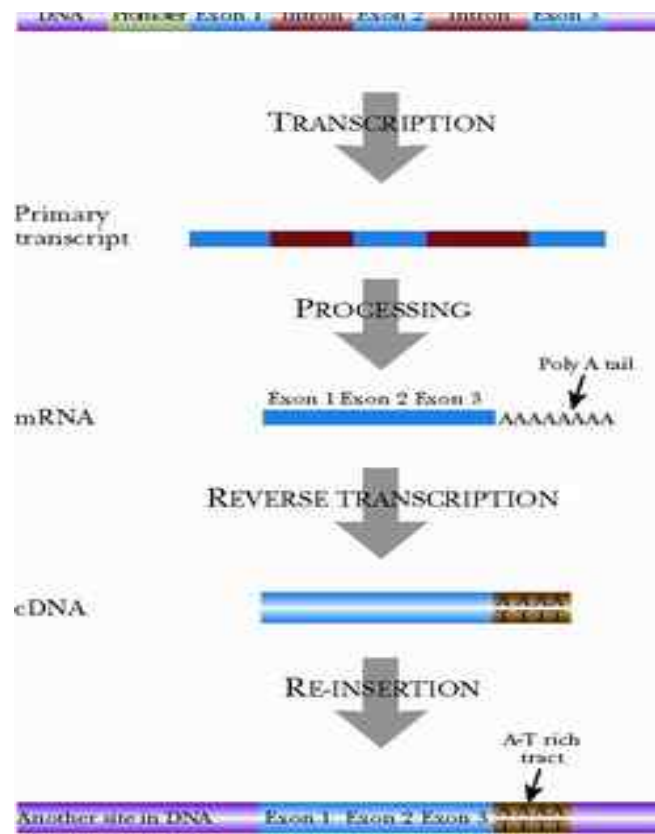
DNA Duplications:

Two phenomenon are common in DNA duplications:

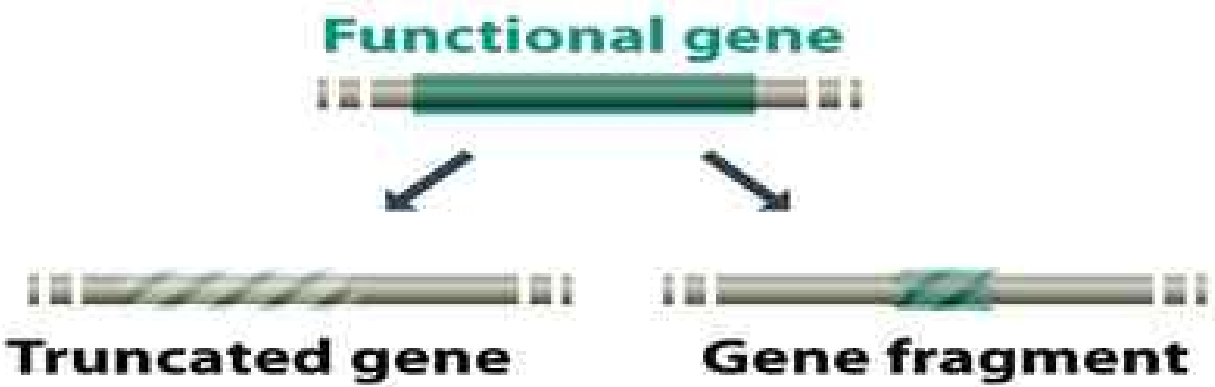
- Slipped strand mispairing
- Unequal crossover during recombination

Pseudo-Genes;





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



Duplicated Genes and Pseudo Genes - Conclusion

- Clustered together in genome.
- Formed by duplication of an ancestral gene
- Nonfunctional copies of genes

52.Repetitive DNA:

Chromosomes - highly dynamic

- 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.

Simple Sequence Repeats

Simple sequence repeats can be;

- 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 these types of tandem repeats are 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 can be;

- Simple sequence of DNA
- DNA transposons
- LTR retrotransposons
- Non LTR retrotransposons like LINE, SINE

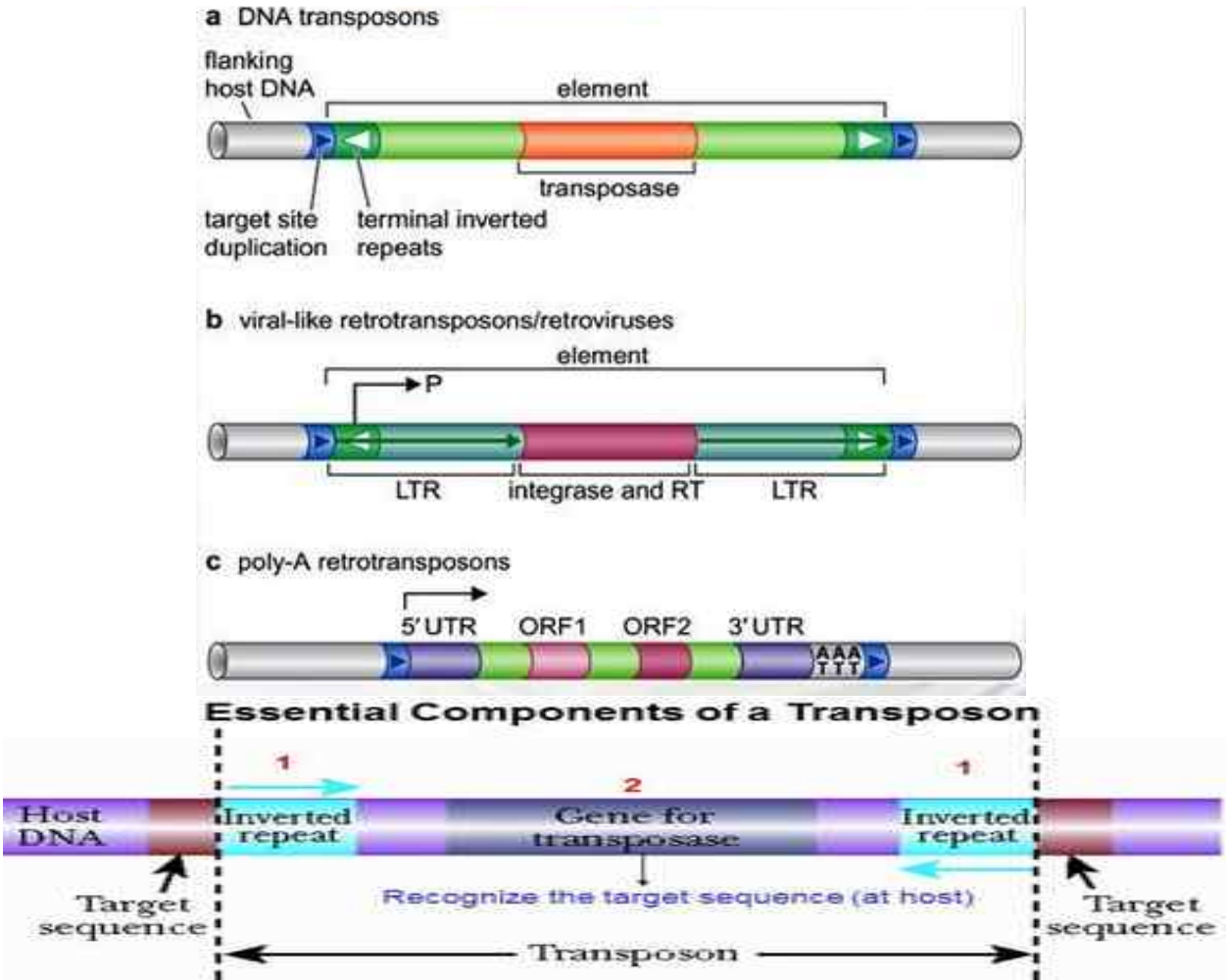
53.Mobile DNA

- It is the type of DNA that move within genomes
- Most of moderately repeated DNA sequences found throughout higher eukaryotic genomes • Most mobile DNA are in the form of transposable elements
- Includes DNA-based transposons and retro-transposons.
- Some encode enzymes that catalyze movement.

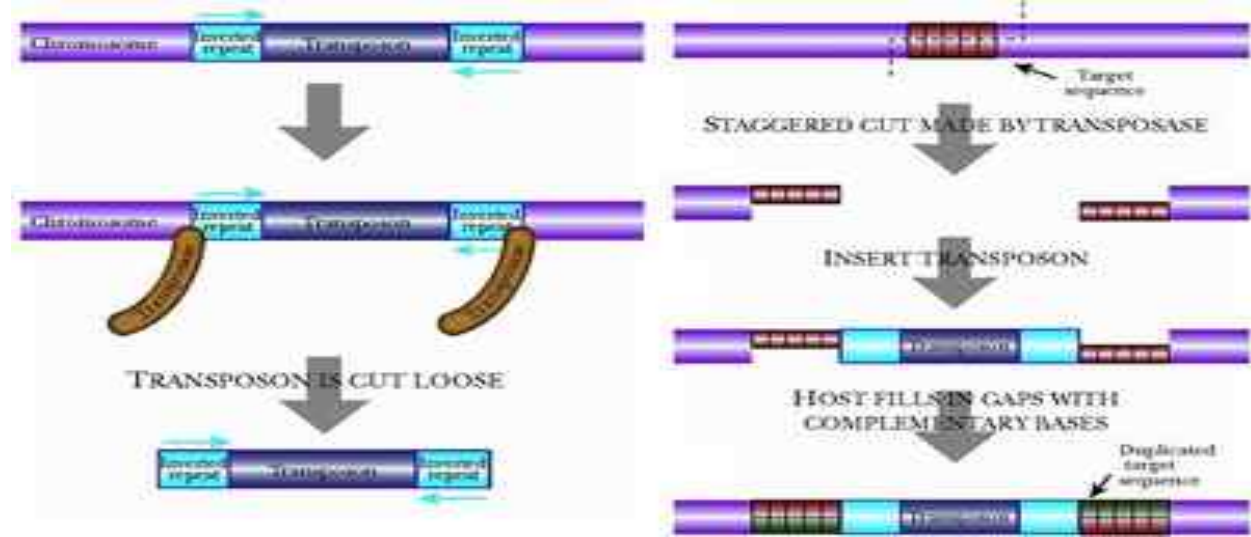
Mobile DNA – Types of Transposons

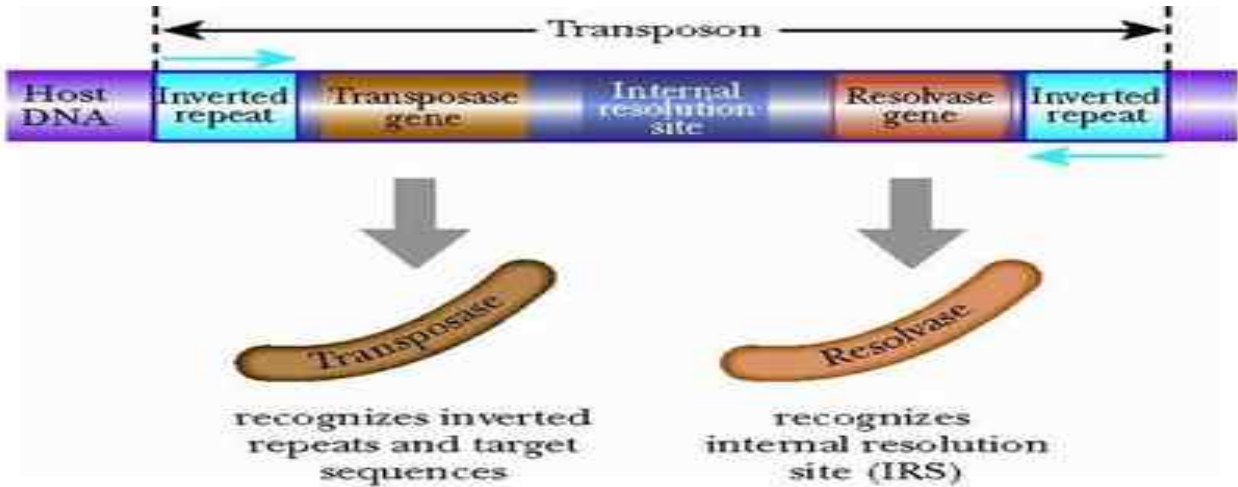
There are following three main types of transposons;

- (a) DNA transposons
- (b) Viral-like retrotransposons/retroviruses
- (c) Poly-A retrotransposons

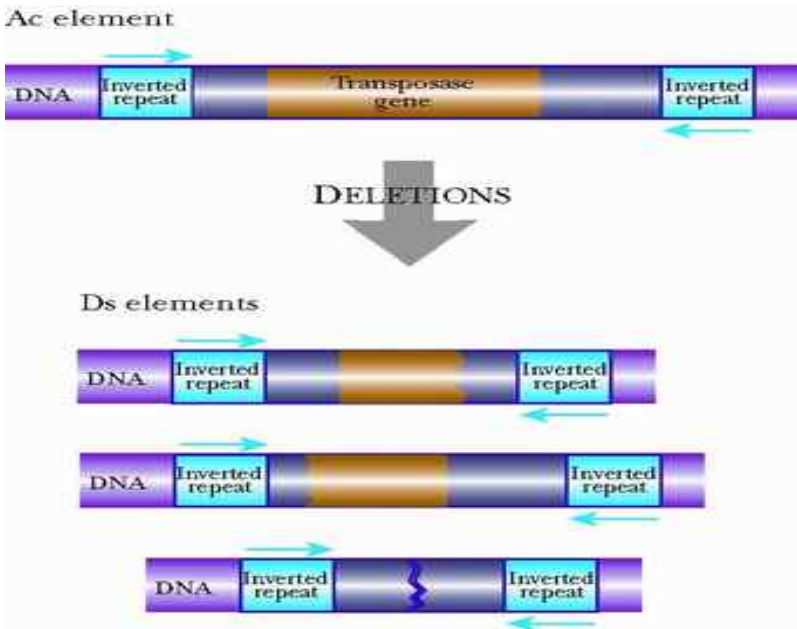


Transposon is cut off by transposase (Enzyme) and inserted on another site

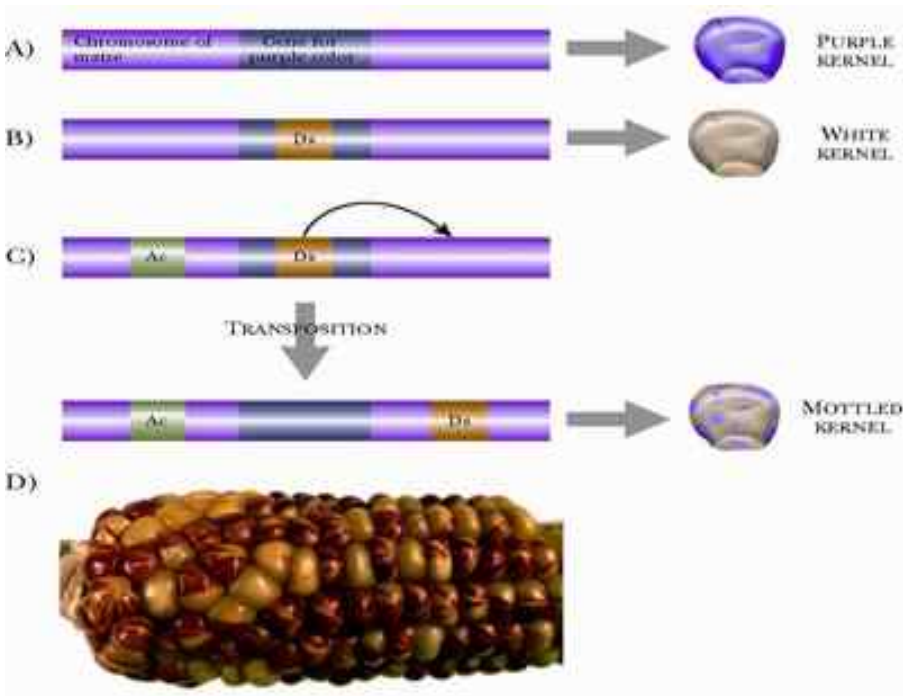




Ac and Ds Elements of Maize – Higher Life Form



Movement of Ds Elements Gives Mottled Corn



Mobile DNA - Transposition

- Transposition is movement of mobile DNA
- Involves copying of mobile DNA element and insertion into new site in genome

Mobile DNA - Retrotransposition

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

Examples;

- 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		1–5 kb	20,000–40,000	21%
SINEs	Nonautonomous		100–300 bp	1,500,000	13%
Retrovirus-like elements	Autonomous		8–11 kb	450,000	8%
	Nonautonomous		1.5–3 kb		
DNA transposons	Autonomous		2–3 kb	300,000	3%
	Nonautonomous		80–3000 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:

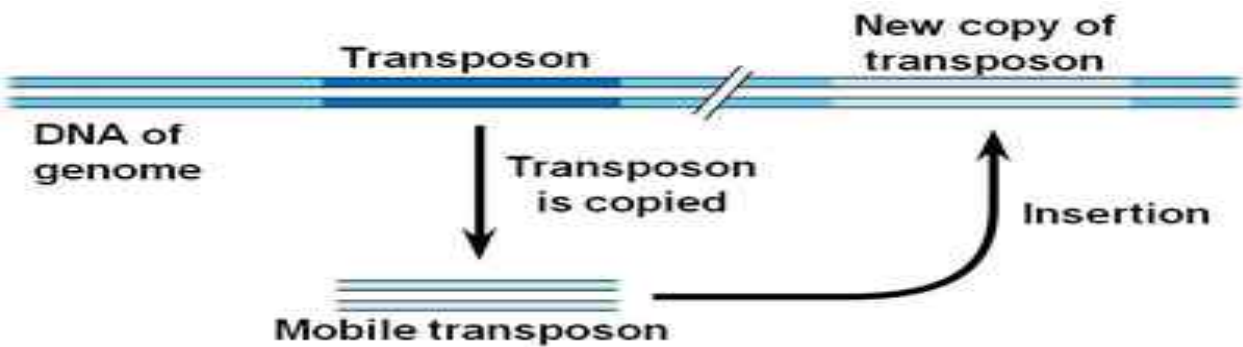
Movement of Transposons

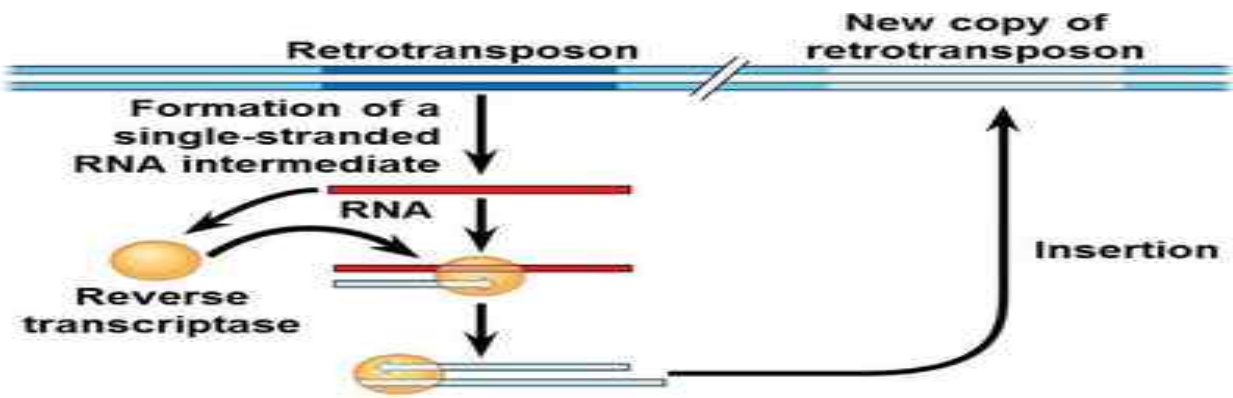
- Eukaryotic transposable elements - two types;
 - 1) **Transposons**, which move by means of a DNA intermediate
 - 2) **Retrotransposons**, which move by means of an RNA intermediate

Transposons

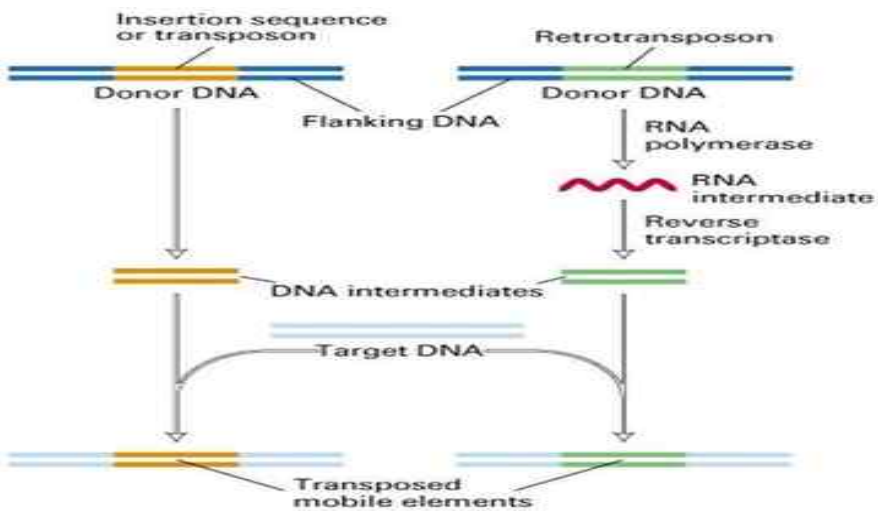
- Transposons - major content of eukaryotic genomes
- ~50% of Human/Mouse
- ~75% of the maize genome
- ~85% of the barley genome
- ~98% of the iris genome

Movement of Transposons (Figure)

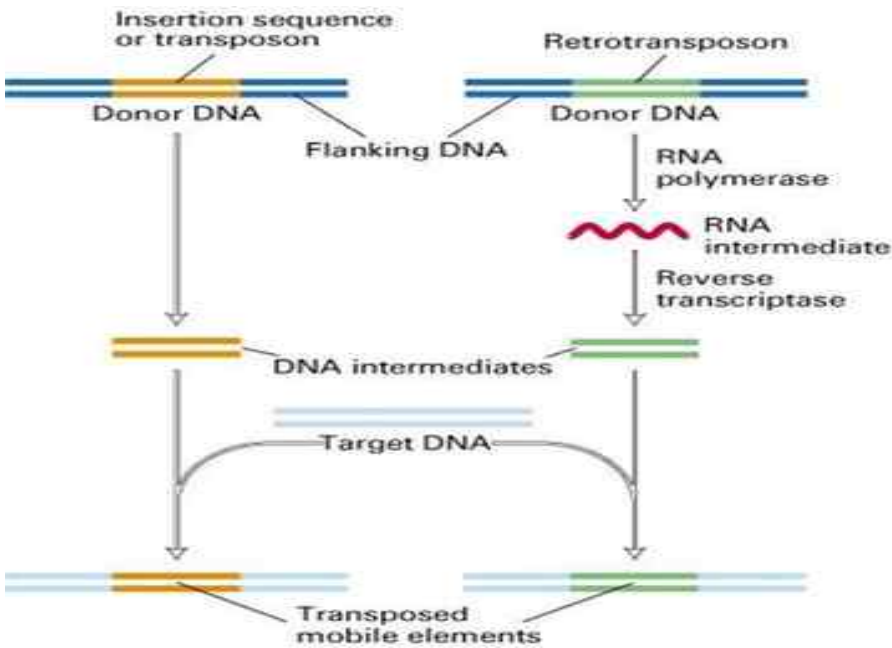




Transposons moves using DNA intermediate (Figure)



Retrotransposon moves using RNA intermediate (Figure)



Conclusion

- Eukaryotic transposable elements - two types;
 - (i) Transposons, which move by means of a DNA intermediate
 - (ii) Retrotransposons, which move by means of an RNA intermediate

55. Transposons in Prokaryotes and Eukaryotes

- Two types of transposons in prokaryotes
i.e. o Insertion Sequence (IS)
 o Transposons (Tn) and Bacteriophage Mu

Transposons in Prokaryotes: 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).

Fig: (a)

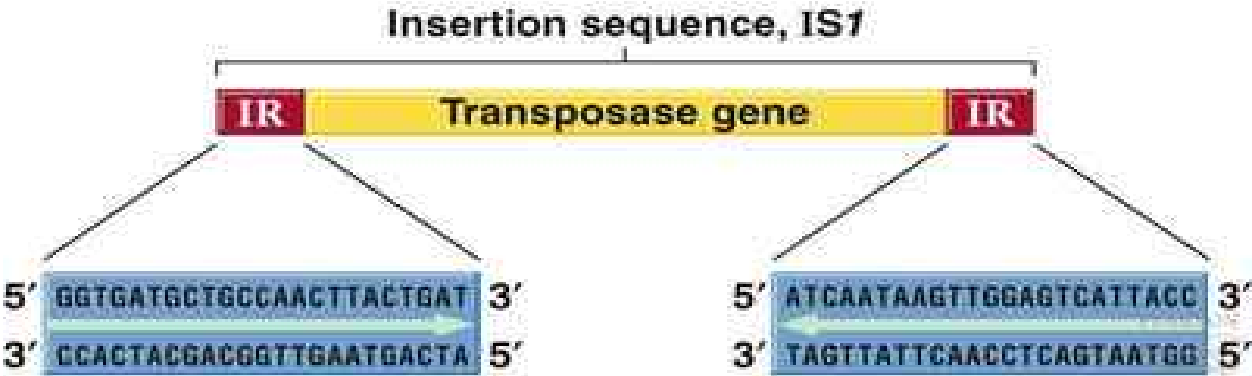
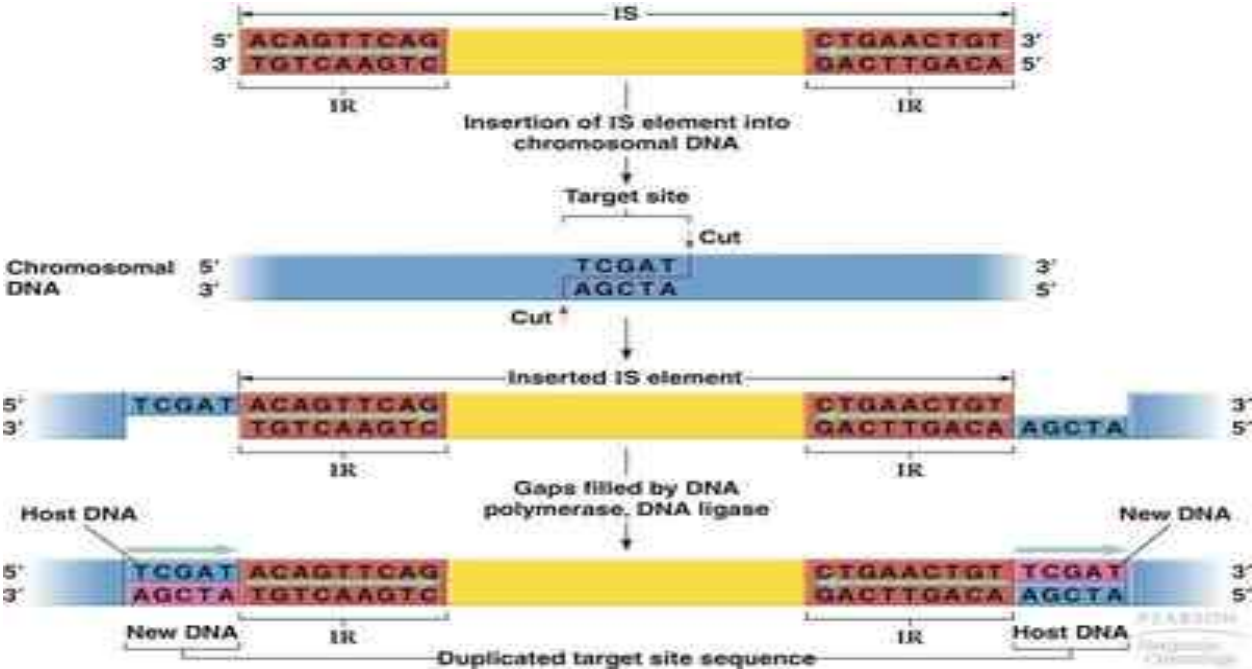


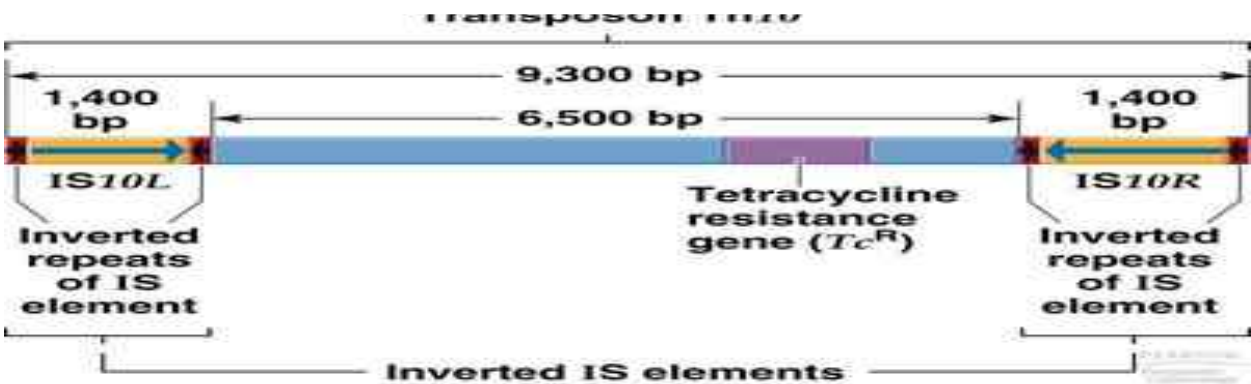
Fig: (b)



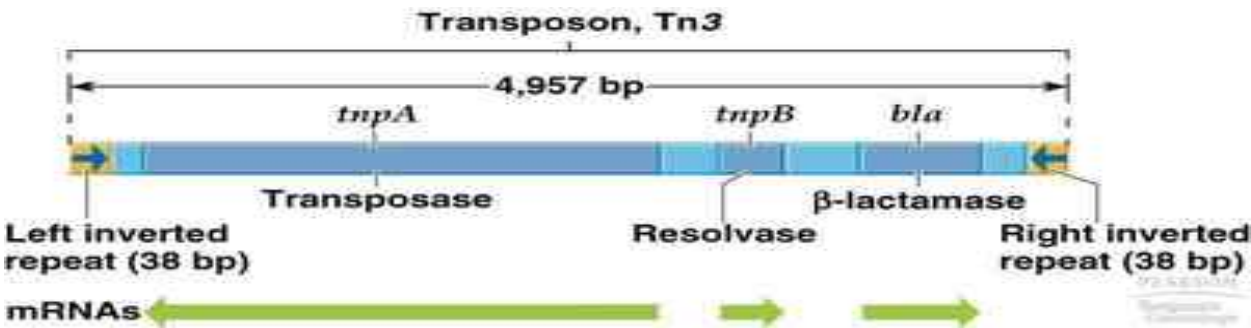
Transposons in Prokaryotes

- 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.

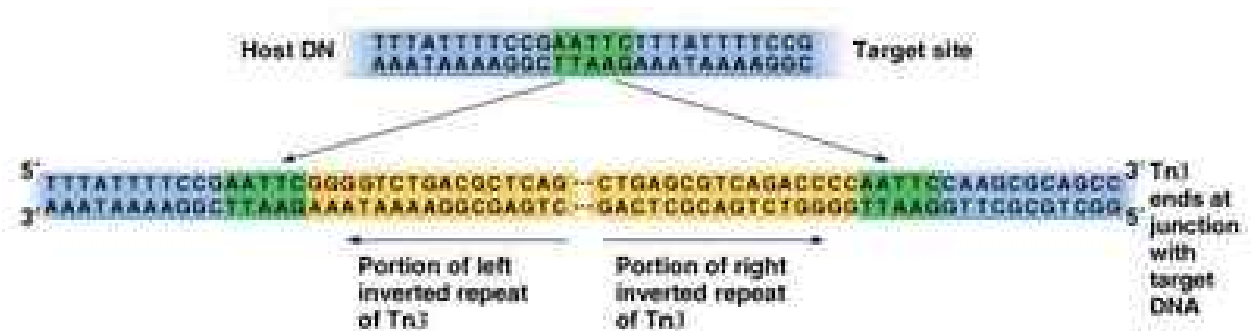
Transposons in Prokaryotes: Composite Tn10



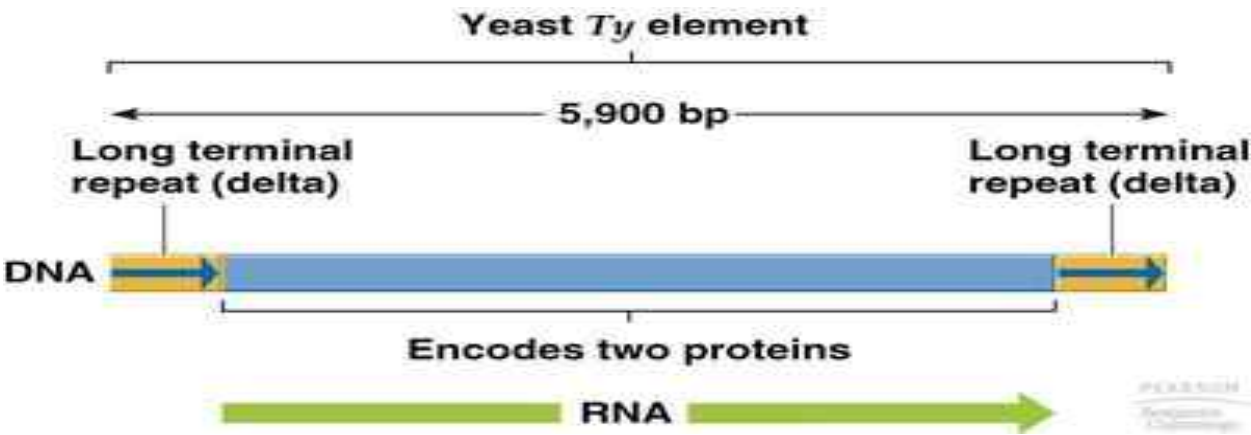
Transposons in Prokaryotes: Non-composite Tn3



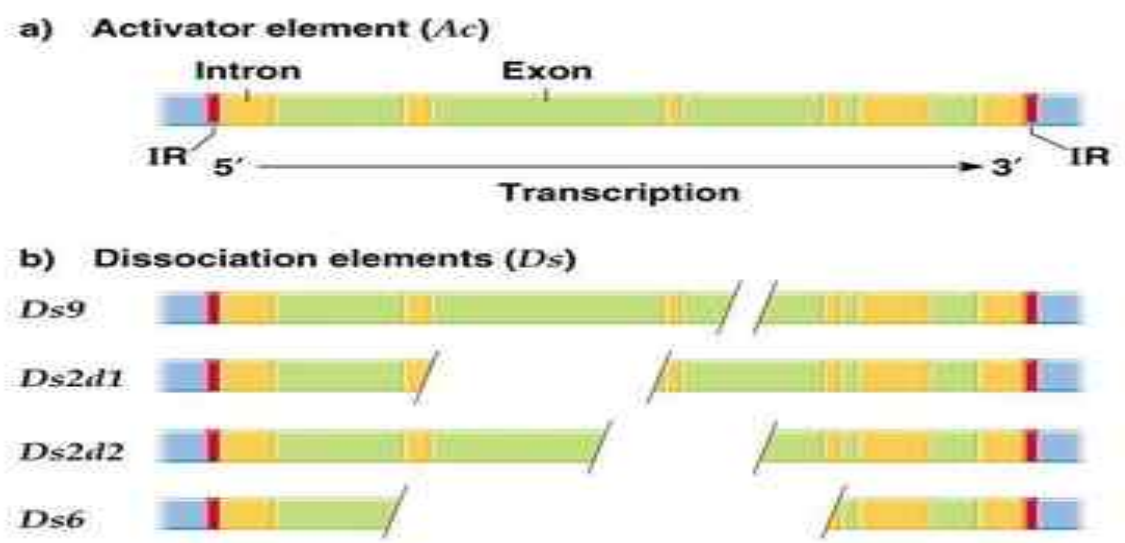
Transposons in Prokaryotes: Non-composite Tn3



Transposons in Eukaryotes: Ty Elements in Yeast



Transposons in Eukaryotes: Ac and Ds Elements



Transposons in Eukaryotes: *Ac* and *Ds* Elements in Maize

- *Ac* is 4,563 bp, with 1 1-bp imperfect terminal IRs and 1 transcription unit producing a 3.5 kb mRNA encoding an 807 amino acid transposase.
- *Ac* activates *Ds* to transpose or break the chromosome where it is inserted.
- *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*

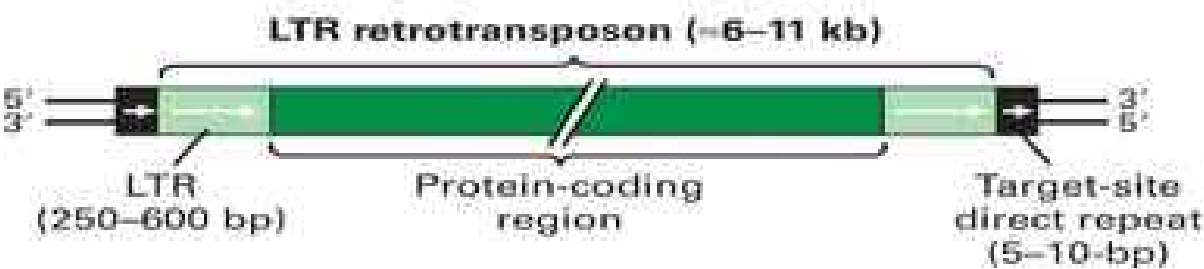
Transposons – Conclusion

- Transposons are found in both Prokaryotes and Eukaryotes.

56.Long Terminal Repeats

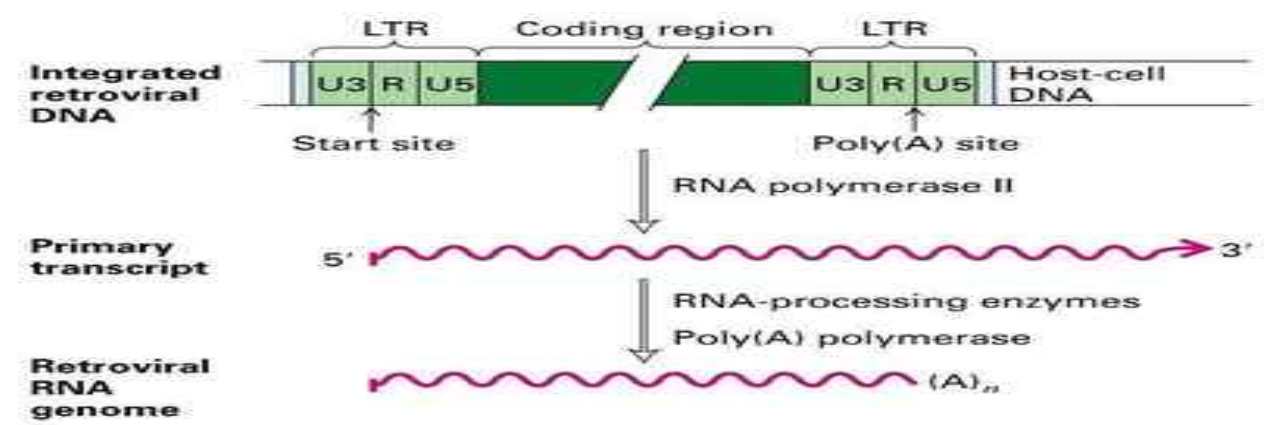
- Eukaryotic retrotransposons fall into two major groups:
 - LTR retrotransposons
 - Non-LTR retrotransposons
- LTRs stand for Long direct Terminal Repeats.
- LTRs consist of 250-600 bp direct repeat sequences located at the ends of the retrotransposon coding region • Flank viral retrotransposons and retroviruses
- Contain regulatory sequences like transcription start site and poly (A) site

Long Terminal Repeats (Figure: a)

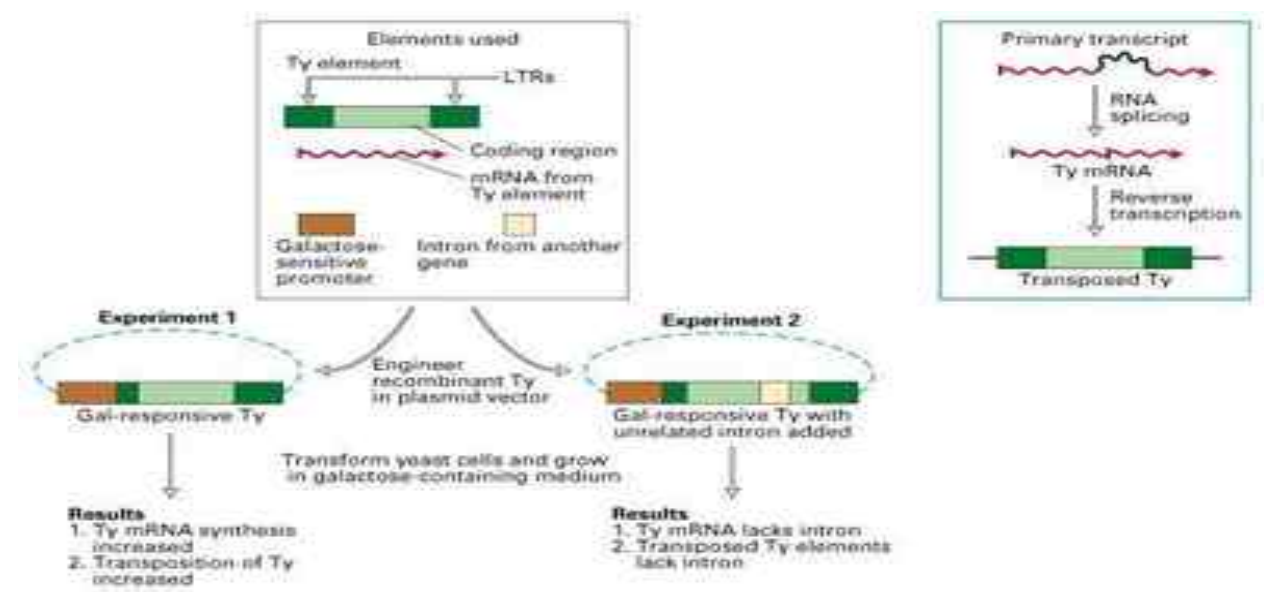


Long Terminal Repeats (Fig: b)





Long Terminal Repeats (Fig: d)

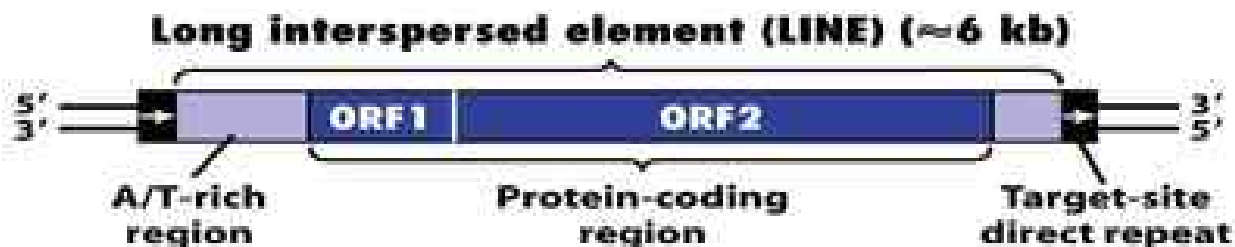


Conclusion

- 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)

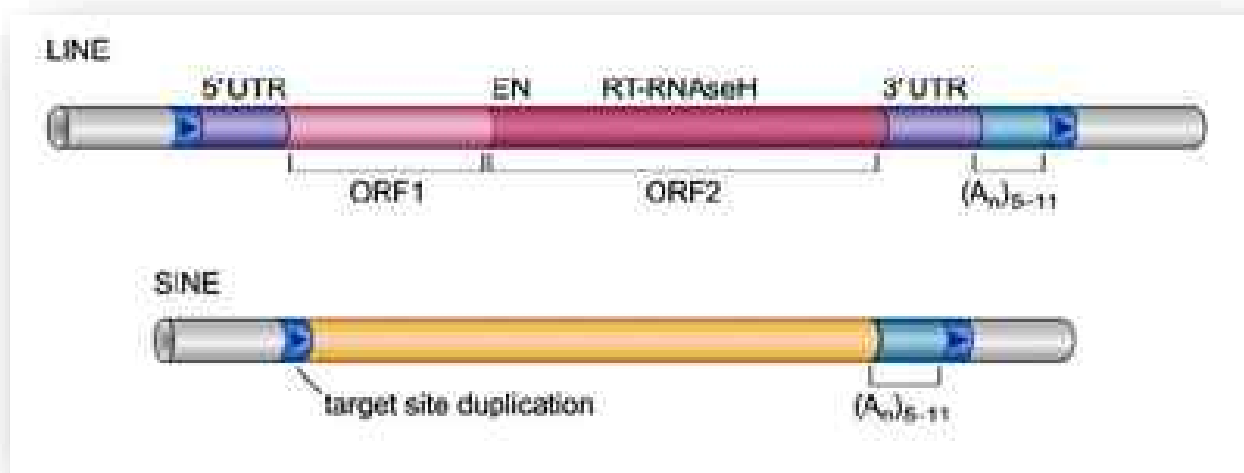
- Non-LTR retrotransposons consist of two sub-types;
 - Long Interspersed Nuclear Elements (LINEs)
 - Short Interspersed Nuclear Elements (SINEs)
 - High copy numbers, in the plants species
 - Widespread in eukaryotic genomes **LINEs**
 - LINEs possess two ORFs, which encode all the functions needed for retrotransposition
- Functions of ORFs include;
- Reverse transcriptase
 - Endonuclease activities, in addition to a nucleic acid-binding property needed to form a ribonucleoprotein
- LINEs are of 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

SINEs

- SINEs use the LINE machinery and function as non-autonomous retro-elements
- ~300 base pairs
- *Alu*
- ~ 11-13% of human genome



The sequences of LINE and SINE look like simple genes. (Figure: LINEs & SINEs)

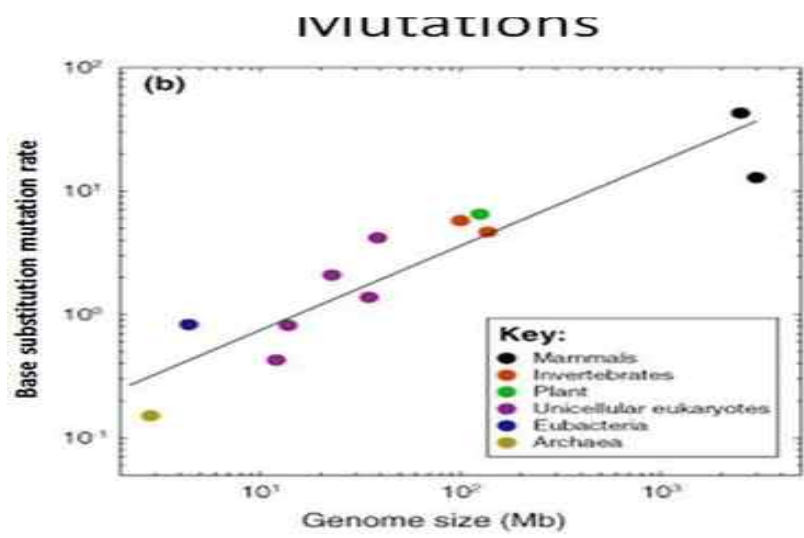
- Short DNA sequences ~ 300 bases that represent reverse-transcribed RNA molecules
 - SINEs do not encode a functional reverse transcriptase protein and rely on other mobile elements for transposition
 - In some cases they may have their own endonuclease that will allow them to cleave their way into the genome
 - 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
- LINEs and SINEs - 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



- Recombinations
(Figure: a)

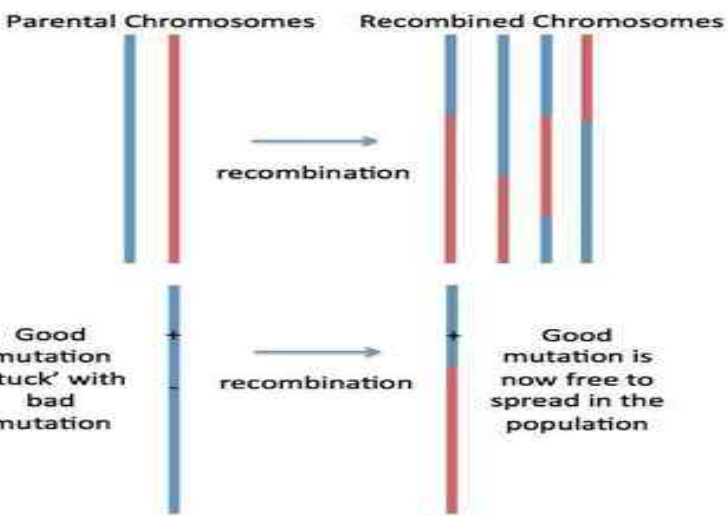
Recombination



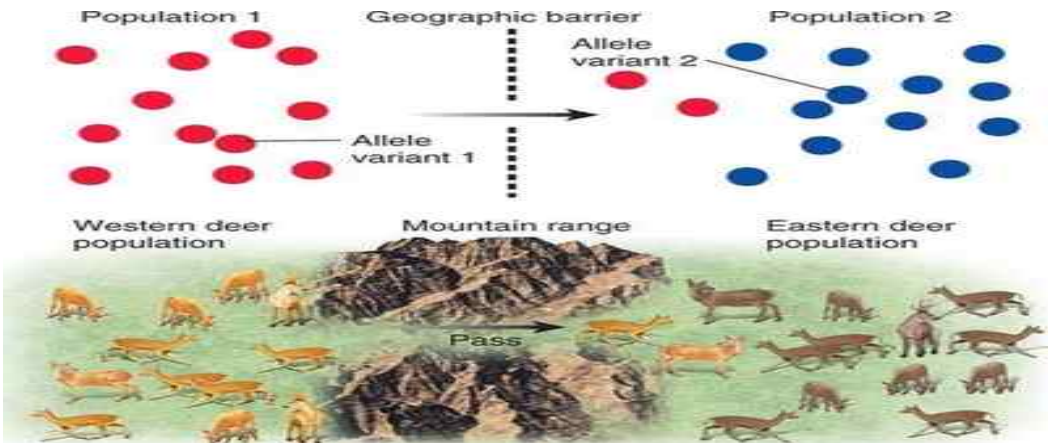
Recombinations

Shuffling gene variants (alleles) in a population

(Figure: b)

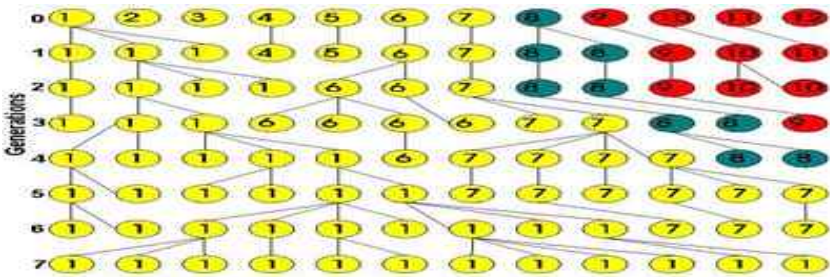


Genes Flow
Movement of alleles from one population to another population.



- Genetic Drift

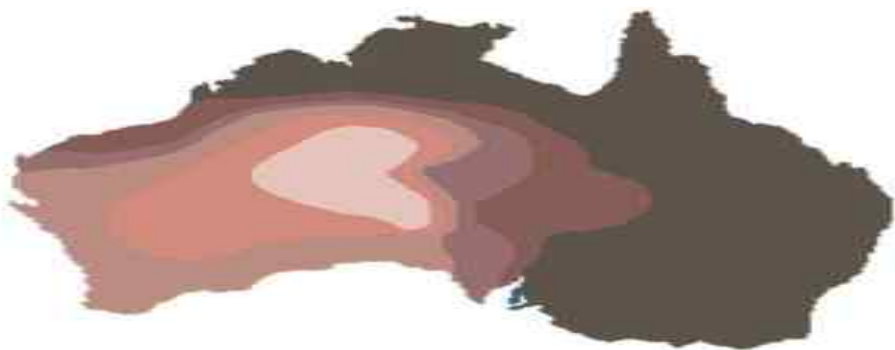
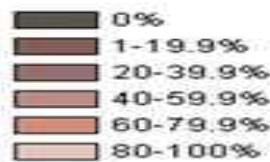
Change in the frequency of a gene variant (allele) in a population due to random sampling of organisms.



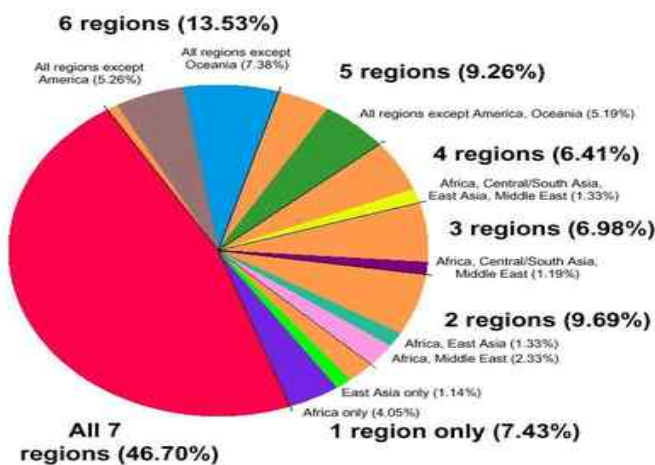
- Humans Skin Color



Frequency of Yellow-brown Hair



- Microsatellites



(377 autosomal microsatellites loci in 1056 individuals from 52 populations of seven regions)

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:

Scientists have worked to detect the foreign DNA. This foreign DNA is our gene of interest. This can be inserted into a suitable replicon, these replicons are the fragments of DNA and they can be used as cloning vector. Different types of cloning vectors are available. Among them the most important are the bacterial plasmids and bacteriophages as well.

Plasmids are the extra chromosomal DNA that has its own region of replication, that's why they are called as replicons. These plasmids can multiply independently without any assistance.

When foreign DNA is inserted in them, they multiply without any assistance of the host chromosome.

- The fragments of DNA can be inserted to a suitable replicons-cloning vehicles
- Plasmids and bacteriophages are most suitable vectors

60. Isolation of genomic DNA from E. coli:

To perform gene manipulation, first step is isolation of genomic DNA. Genomic DNA represents whole gene complement of the organism. It also includes our gene of interest.

To isolate DNA (For example if we want to proceed with E.Coli, that is a microbe), then microbes (E. Coli) will be grown in a suitable culture medium. Then they will be incubated for 24 hours. After 24 hours visible mass of E.Coli cells (or microbial cells) will be produced and then the cells will be separated or harvested by centrifugation. Now this process will be further proceed for the isolation of genomic DNA.

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

61. Agarose gel electrophoresis:

After DNA isolation we can analyze or visualize them by using Gel electrophoresis.

- **Agarose** is convenient for separating DNA fragments ranging in size from a few hundred base pairs to about 20 kb. **What is Agarose?**

Agarose is extract of sea weed or Algae grown in marine environment or ocean. After extraction it is converted into powder form. This Agarose can be dissolved in suitable buffer medium and then it is melted. Then it is allowed to be solidified. After solidification DNA sample can be loaded on it.

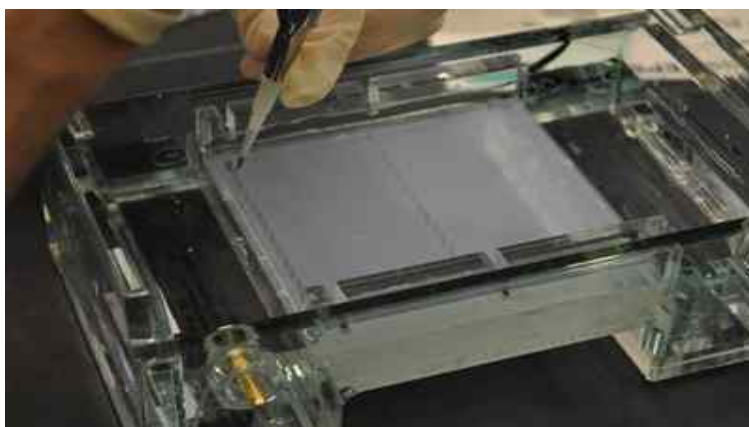
When we are using Agarose to separate DNA fragments, it is very important to adjust its concentration. For example if we want to analyze small DNA fragments, higher concentration (2%, 3%, 4%) of Agarose is used and if we want to analyze large DNA fragments then low concentration (0.7-1%) of Agarose is used. Because if we use low concentration, pore size of gel will be large that will allow larger molecules to pass through this gel. If we use high concentration then pore size will reduce and only smaller DNA fragments can pass through this gel.

What are different factors that can influence the mobility of DNA of Agarose gel electrophoresis?

- Concentration of gel either high concentration or lower concentration.
- Ionic strength of buffer.
- Voltage
- Ethidium bromide (A chemical used as a stain to dye gel)

Polyacrlamide gel electrophoresis

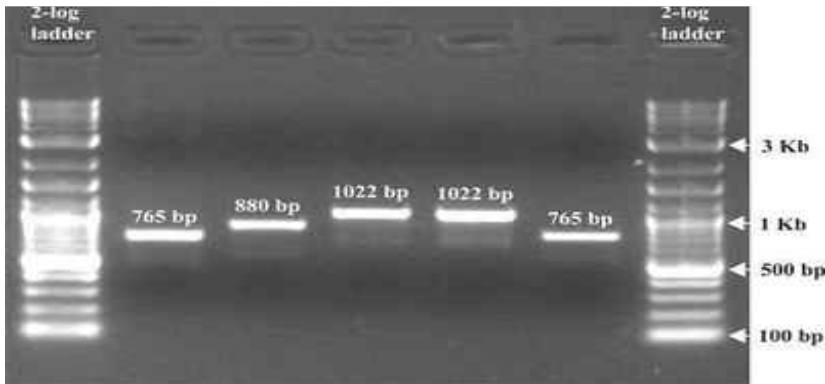
- Polyacrlamide is preferred for smaller DNA fragments ▪ It is also used to isolate proteins molecules



Gel electrophoresis apparatus



Visualization of gel under UV light



Visualization of gel on gel doc system

62.Nucleic acid blotting

- Blotting describes the immobilization of sample nucleic acids or proteins on to a solid support, generally nylon or nitrocellulose membranes are used for this purpose.
- 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

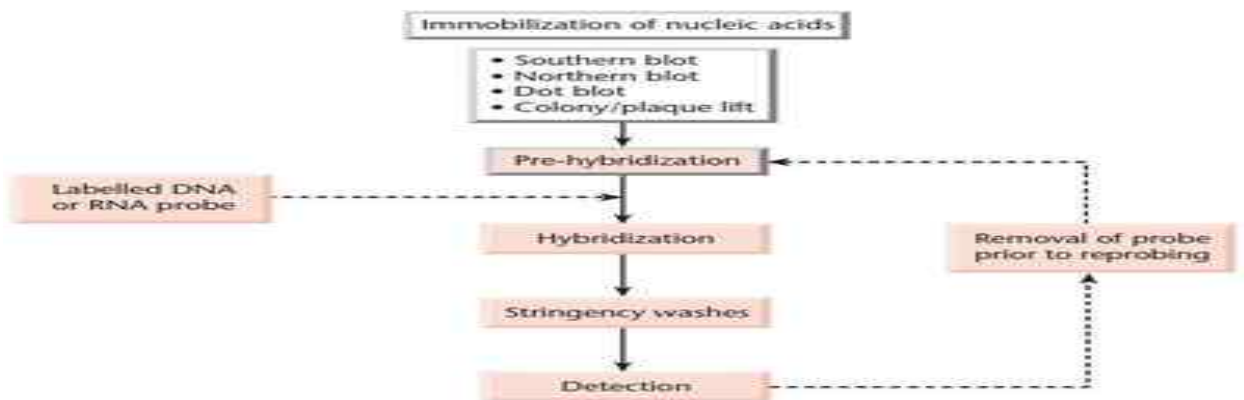


Fig: Nucleic acid blotting and hybridization

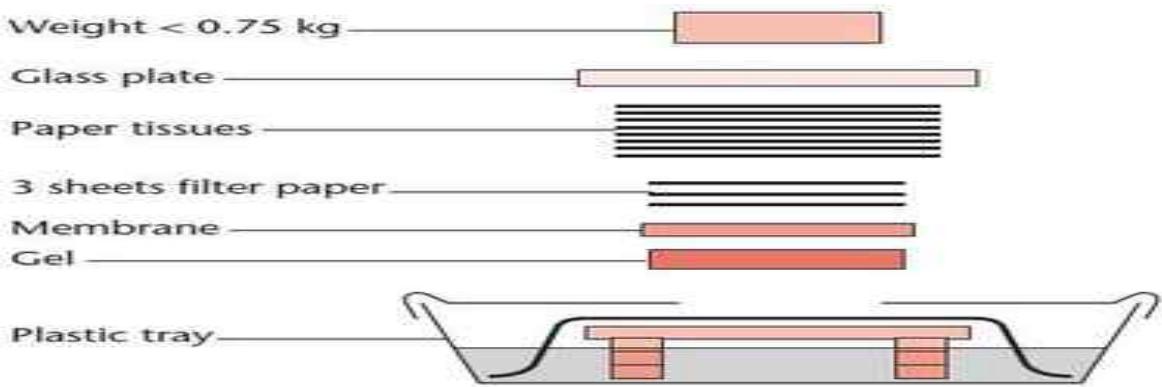


Fig: A typical capillary blotting apparatus

63.Autoradiography

electrophoresis

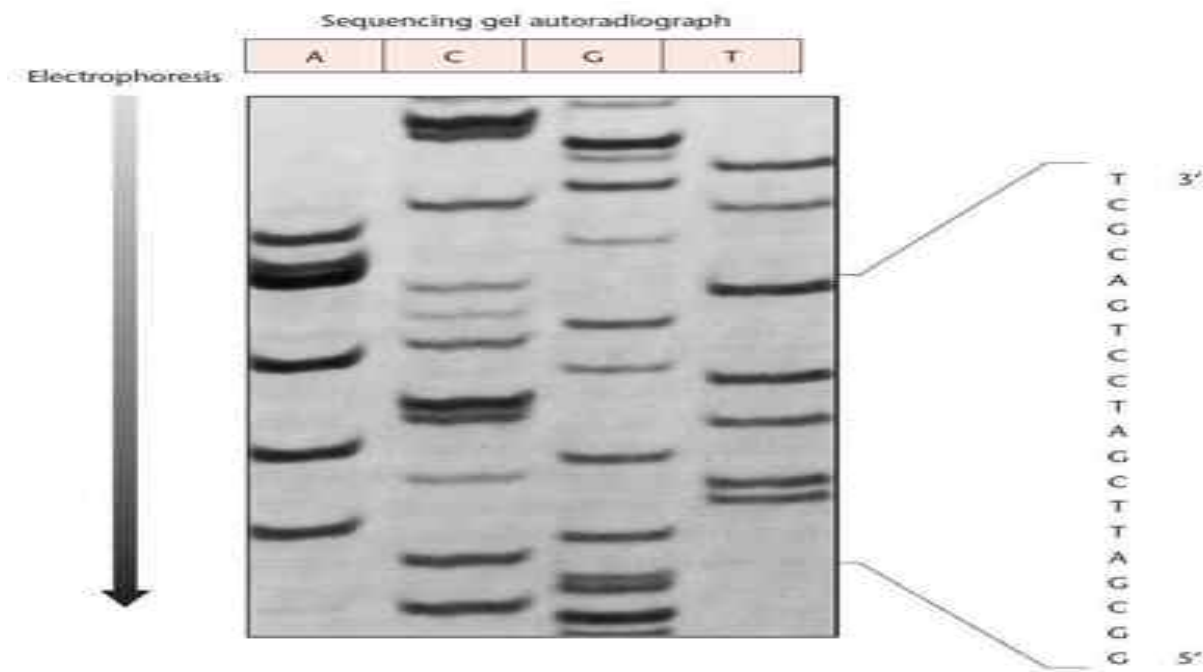


Fig: Autoradiograph of sequencing gel

64.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 labeled probe. Probe is labeled with radioactive substances like Phosphorus 32.
- After hybridization reaction, membrane is washed and hybridization are detected by autoradiography

Fig: Overview of Southern blotting

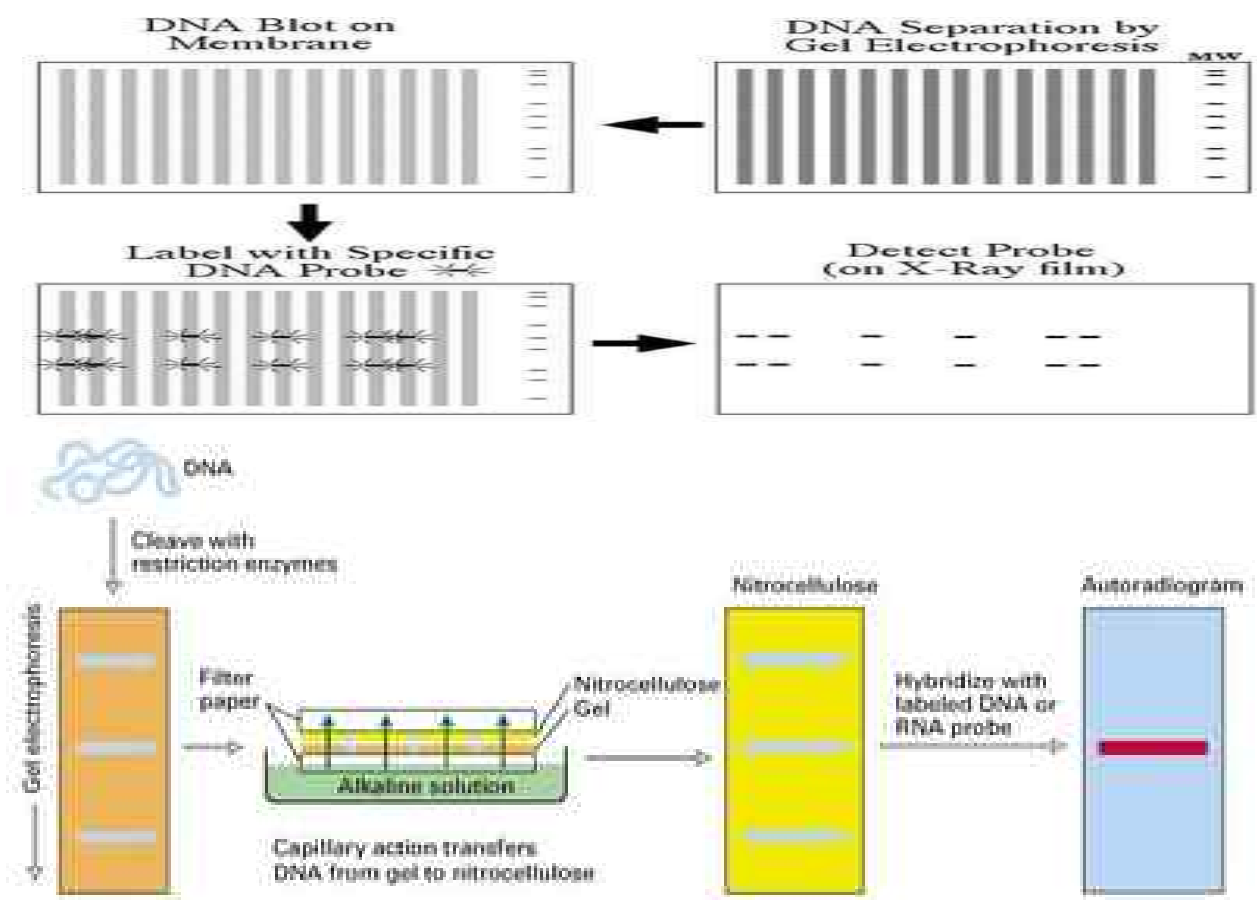
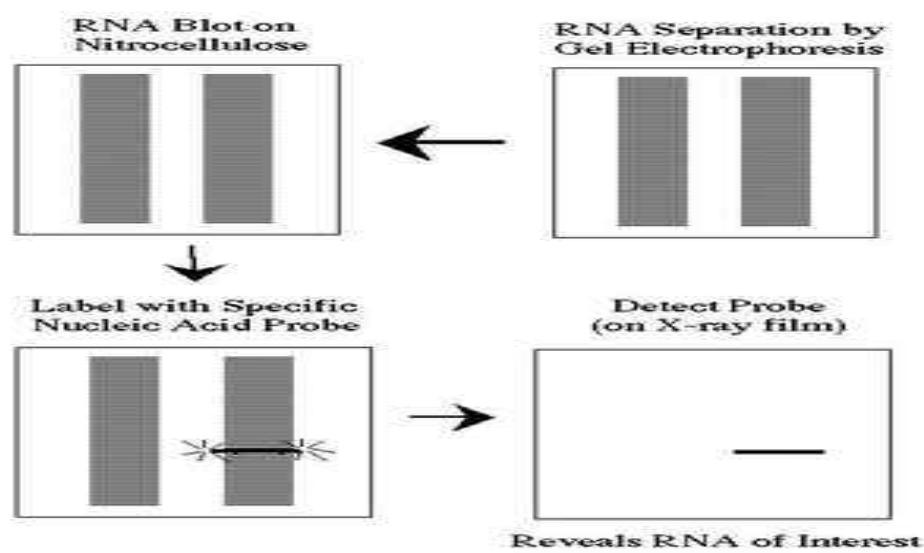


Fig: Southern blot hybridization

65.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



Northern

Fig: Overview of blotting

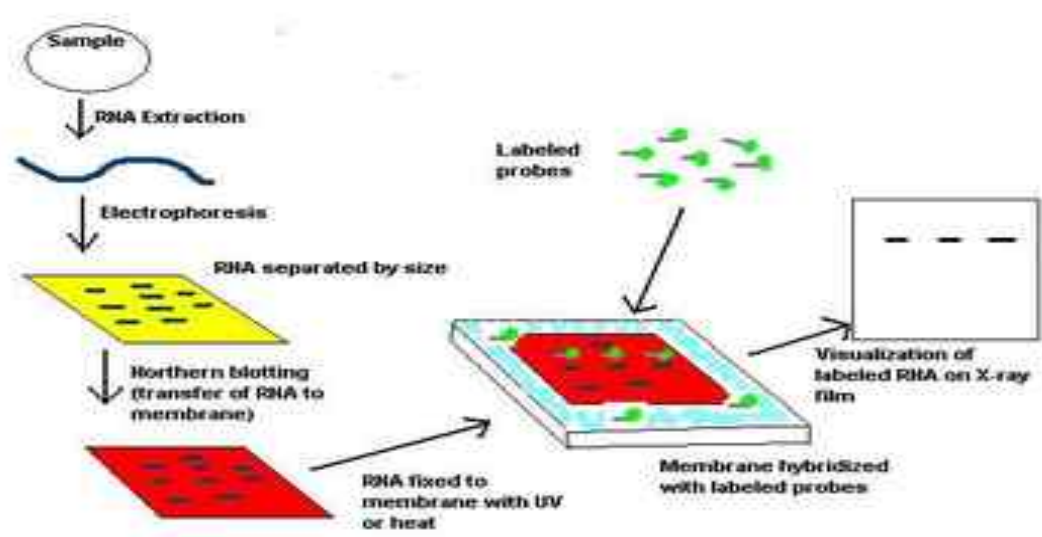


Fig: RNA detection by Northern blotting

66.Western blotting

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

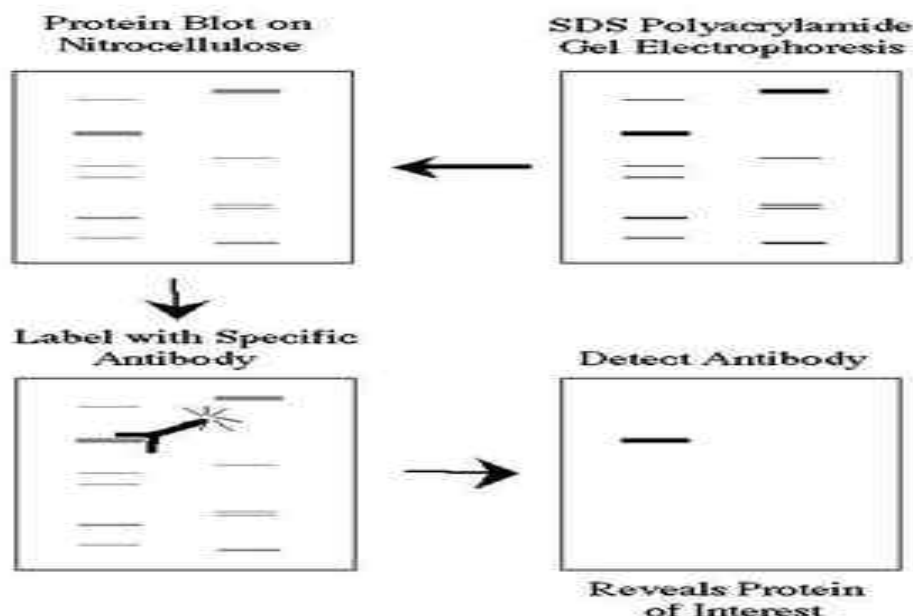


Fig: Western blot method

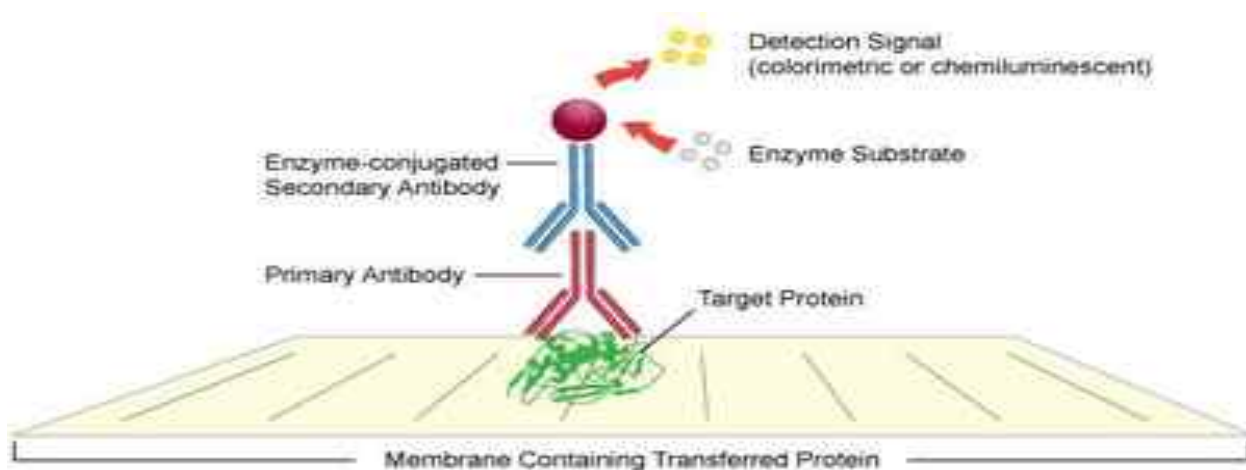


Fig: Detection in Western blot

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 (Ability of the cell to take foreign DNA)

Explanation: Transformation is the genetic alteration of the cell by uptake of foreign DNA from its environment. If we consider the example of *E. coli* as a host organism; we can say that it is the uptake of recombinant DNA like vector or plasmid by the recipient *E. coli* cell from a mixture of plasmids. Earlier different workers attempted to transform *E. coli* but failed because of the following reasons: Microbes produce different types of restriction enzymes. These enzymes continuously monitor incoming foreign DNA and if cell realize that it is a foreign DNA, it is being attacked and degraded by the restriction enzymes of the cell. So restriction enzymes are the defense mechanism of the cell.

Other reason was the competency of the cell. “**Competence** means that what is the ability of the cell to take the foreign DNA”.

In 1928 Fredrick Griffith tried to transform cells and he was successful to some extent. He was able to transform non-pathogenic strains of streptococcus pneumonia by transferring the genetic material from pathogenic strains of streptococcus pneumonia. He mixed non-pathogenic strains with the solution of heat killed streptococcus pneumonia and then he realized that something transferred from pathogenic heat killed cells to non-pathogenic cells that transformed non-pathogenic cells into virulent strains. Griffith that time didn't know about

DNA.

In 1970 a few workers were able to transform E.Coli in the presence of Calcium Chloride and they transformed it with the help of Bacteriophage lambda DNA.

In 1972 E.Coli was transformed with the plasmid DNA by scientists.

Factors that can influence transformation efficiency:

1. Concentration of chemical that is used i.e. Calcium chloride
2. Temperature

Now it is suggested that cells can be incubated at low temperature with calcium chloride in the presence of gene of interest or foreign DNA and from the low temperature the temperature is suddenly raised to high like 42 to 45 °C that can result in uptake of foreign DNA.

Cells are taken from log phase and mixed with Calcium chloride solution. (Log phase is the phase of bacterial growth where cells are dividing actively/exponentially). So bacterial population doubling after a particular time period. We can take those actively dividing cells and then it can be incubated in ice cold Calcium chloride.

Concentration of Calcium chloride may be 50 mili molar per litter. Then the foreign DNA like (0.1 gram target DNA) can also be incubated with the cells. After 30 minutes of incubation temperature is being raised to 42 or 45 °C so that cells can uptake DNA.

What is the role of Calcium chloride in the solution?

- It can bind with the cell wall.
- It can change the characteristics of bacterial cell wall.
- It can also bind with the DNA.

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

68.Transformation with other organisms

- *E. coli* often remains the host of choice for cloning
- *Bacillus subtilis* (preferable choice because it is natural competent) is a particularly attractive alternative cloning host
- Animal cells, protoplasts (without cell wall) of yeast and plant are susceptible to transformation by liposomes
- Other artificial method for transformation include gene gun
- Lipofaction is a technique in which we use liposomes to transform protoplasts. Liposomes are vesicles made up of phospholipids and are positively charged. They are positively charged because mainly they are made from cationic lipids. They have the ability to enclose negatively charged DNA. After encapsulation liposomes fuse with the cytoplasmic membrane and then it can deliver DNA from the vesicle to the target cell. So lipofaction is also used to transform different type of host cells.

69.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

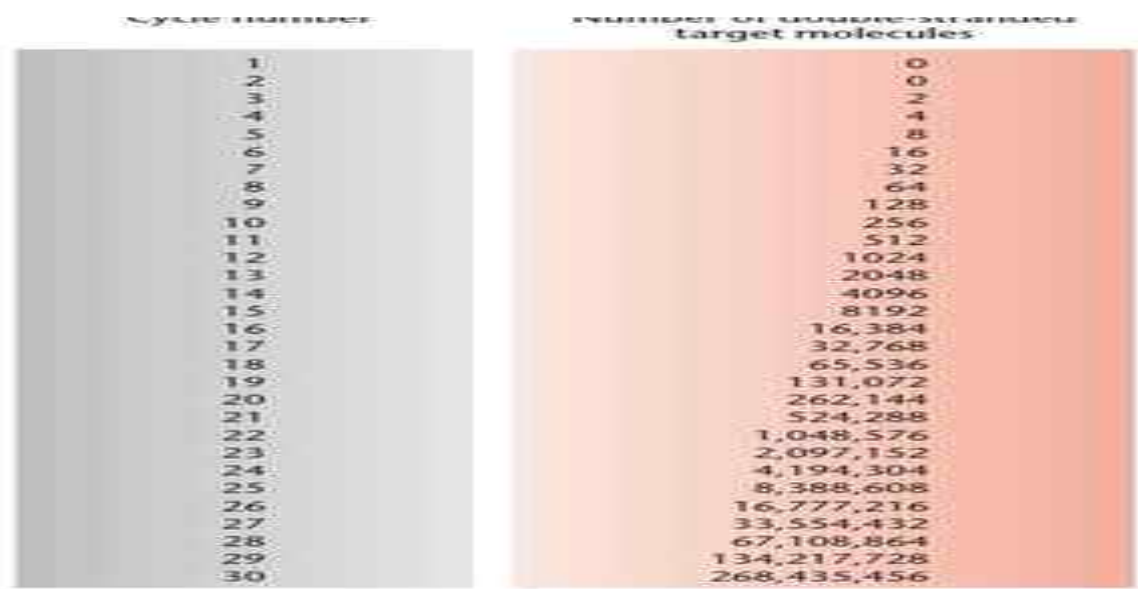


Fig: Theoretical PCR amplification of a target fragment

Applications of PCR:

PCR can be applied for prenatal diagnosis and in forensic science.

70.Cutting and Joining DNA molecules

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 (Efficiency of plating means the number of plaques that are produced during the lysis) 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

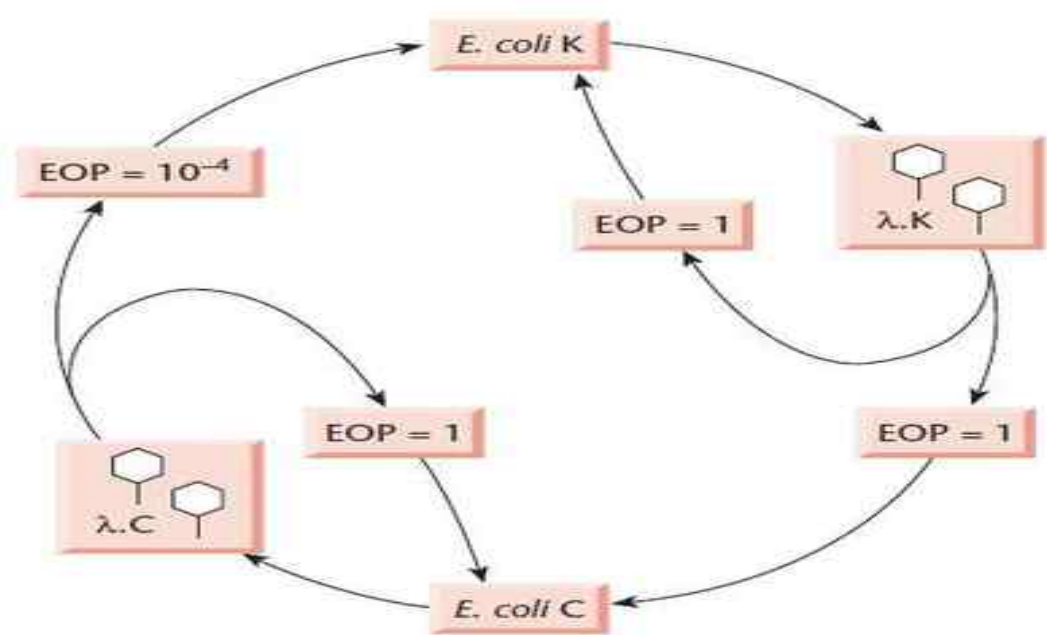


Fig: Host-controlled restriction and modification of phage λ

- 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

72.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

Practically Type I restriction enzyme is not good because it cleaves the DNA randomly. If we analyze the structure of Type I restriction enzyme; it is a multi-protein complex. For the proper functioning of Type I restriction enzyme it requires some Co-factors like for modification that includes the process of methylation of different bases and as well as cutting it requires ATP and presence of S. Adenosyl methyoline

As we know same enzyme is responsible for cutting and methylation. If DNA is un modified, it is being fragmented and once it has been modified by the same enzyme then cutting is not possible. This is one of the basic reasons. Another reason is that they can recognize particular DNA sequence but cut a very from recognition site and cleave the DNA randomly. That's why they are not valuable in 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
- They are more valuable as they recognize particular target sequence
- It use separate enzyme for restriction and modification
- It does not require Co factors to cut the DNA
- Now a days they are commonly used for cutting of DNA
- They cut within the DNA, so they are called 'endonucleases'

Type III

- Type III enzymes have symmetrical recognition sequences but otherwise resemble type I systems and are of little value
- Although they can recognize the specific DNA sequence, they cut away from the recognition site. So DNA fragments that are generated, they are random. So desirable fragment length is sometimes not possible. That's why they are not used in gene manipulation 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

After the discovery of large number of Type II endonucleases, scientists proposed that there must be a naming mechanism which we can call as the nomenclature of the restriction endonucleases.

- 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

the system name, for example, in case of *H. influenzae*

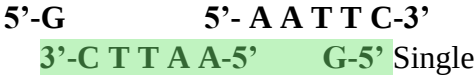
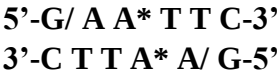
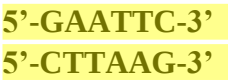
- *R. HindIII* or *M. HindIII*

Examples of restriction endonuclease nomenclature;

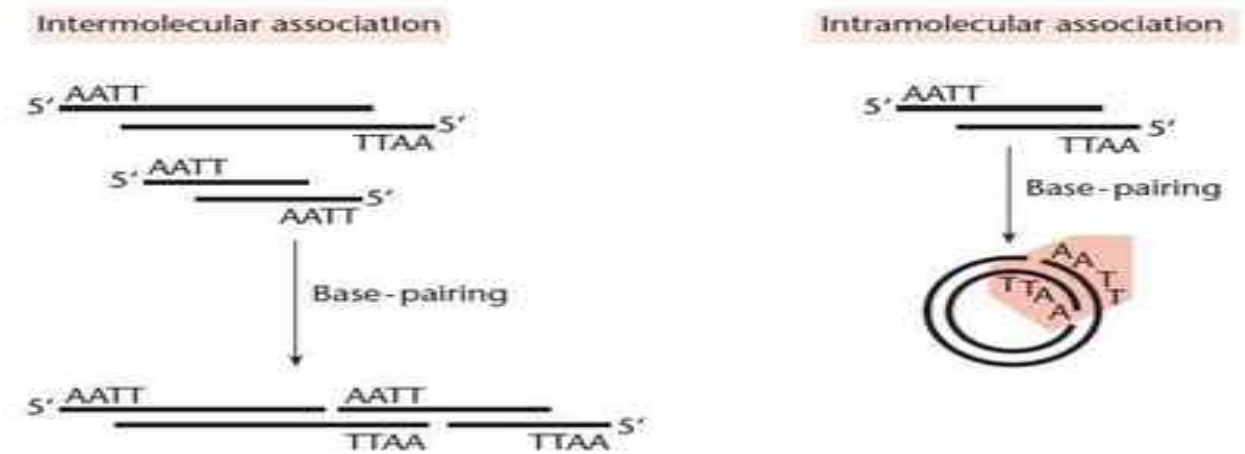
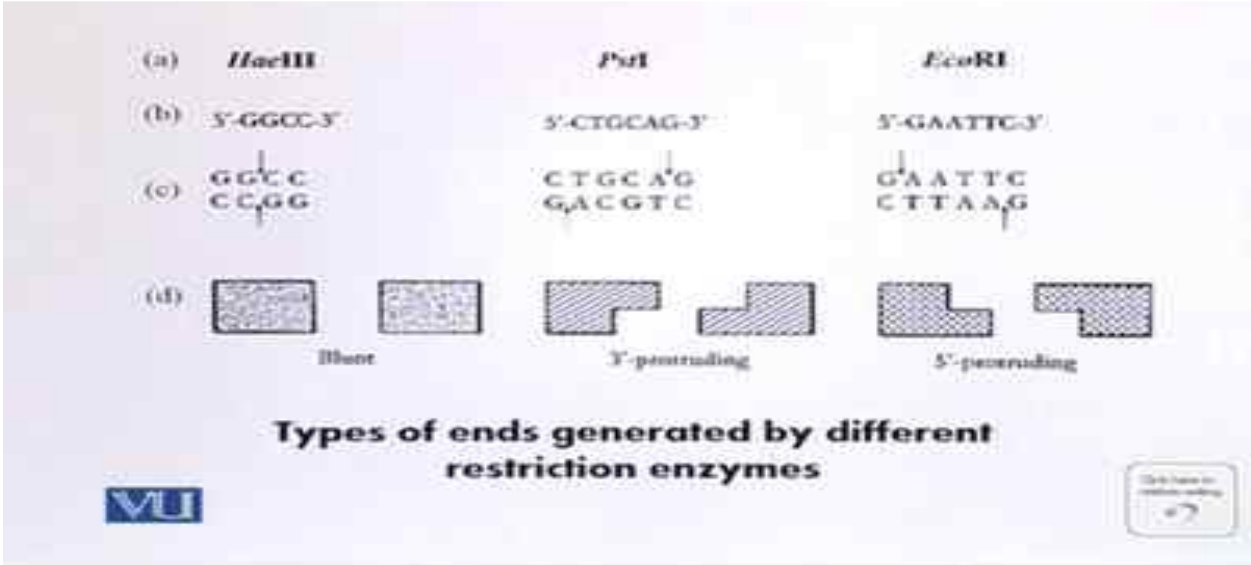
Enzyme	Enzyme source	Recognition sequence
<i>SmaI</i>	<i>Serratia marcescens</i> , 1st enzyme	CCCGGG
<i>HaeIII</i>	<i>Haemophilus aegyptius</i> , 3rd enzyme	GGCC
<i>HindII</i>	<i>H. influenzae</i> , strain d, 2nd enzyme	GTPyPuAC
<i>HindIII</i>	<i>H. influenzae</i> , strain d, 3rd enzyme	AAGCTT
<i>HamHI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

74.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* (words that are read similar either from the right or from the left).



stranded breaks by *EcoR1*



Cohesive fragments of DNA produced by digestion with *EcoRI*

75.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
- It will also be determined by the GC content of an organism
- 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>ApaI</i>	GGGCCC	25000	15000	2000
<i>BamHI</i>	GGATCC	6000	5000	5000
<i>SpeI</i>	ACTAGT	8000	60000	10000

The fragment length that is generated by different restriction enzymes may be different. It not only depends on the length of the recognition sequence, but also depends on the time of incubation with a particular restriction enzyme. If we incubate a target DNA with particular restriction enzyme for longer period of time, it means that we gave more time for digestion and it will generate smaller DNA fragments. If we will reduce incubation

incubation time. So length of fragments that will be generated by partial digestion they may be longer with the same restriction enzyme. So same restriction enzyme is producing shorter DNA fragments as well as larger DNA fragments.

76.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 (Degrade DNA at the terminals, either at the 5' or 3' terminal. They start the degradation of DNA at the ends.)

ii). Endonucleases (Target the DNA from internal sites)

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

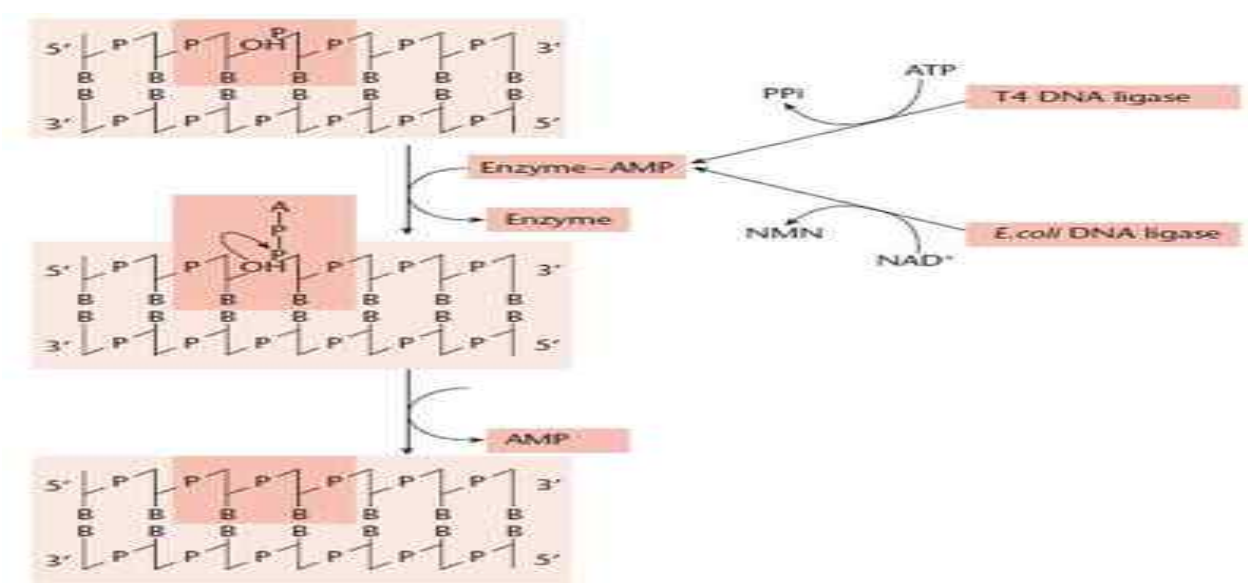
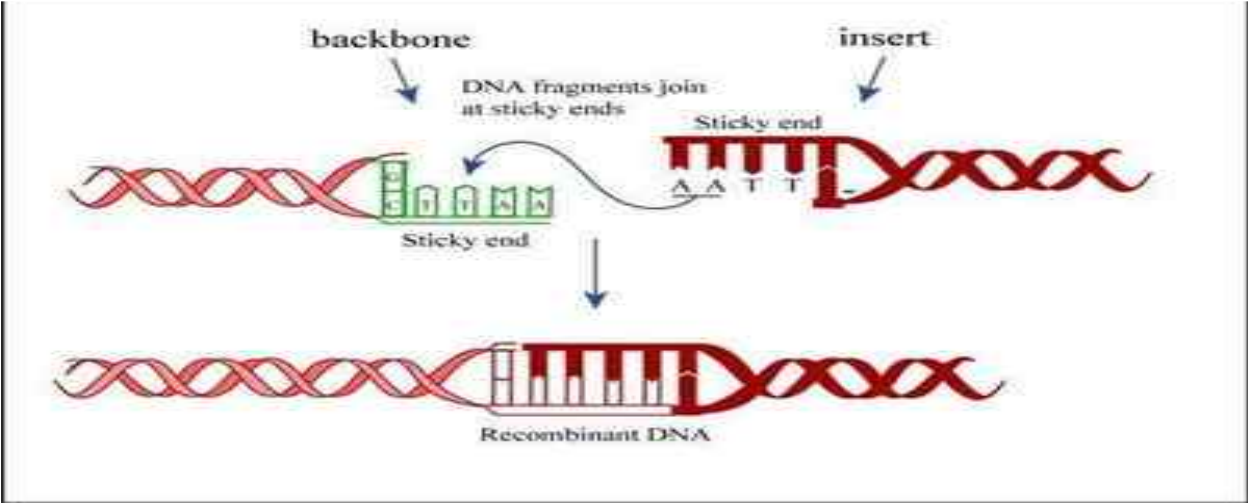
78.Methods of joining DNA fragments

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 deoxynucleotidyl transferase 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

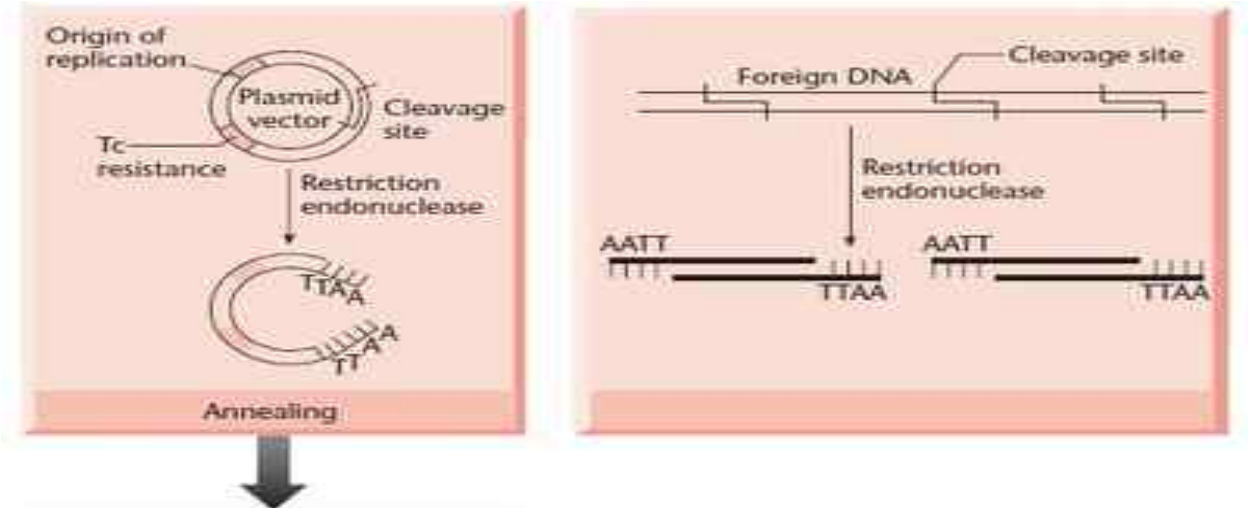
T4 DNA ligase requires ATP while bacterial DNA ligase requires NAD⁺ as a cofactor.

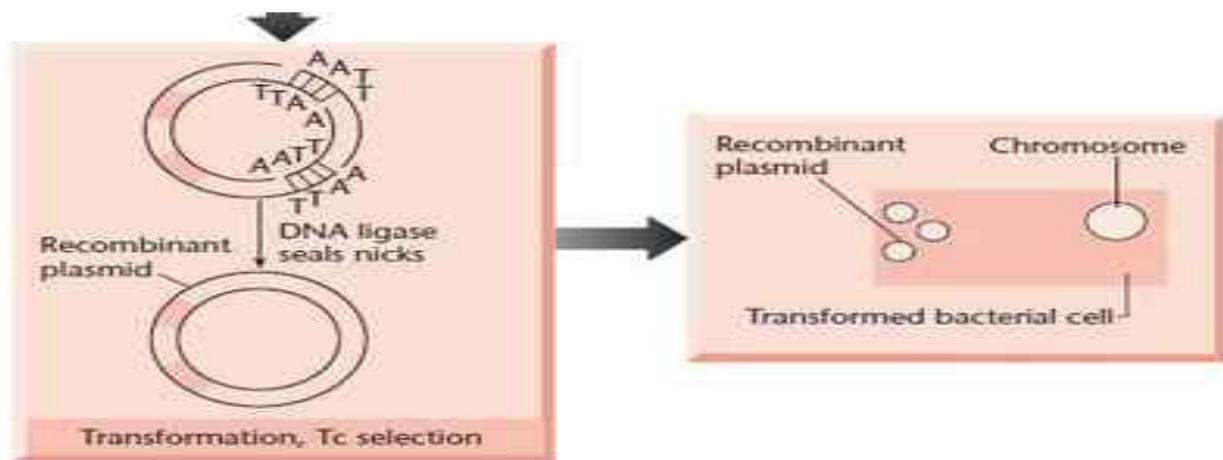


79.DNA ligase to create covalent recombinant DNA

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

Use of DNA ligase to create a covalent DNA recombinant;

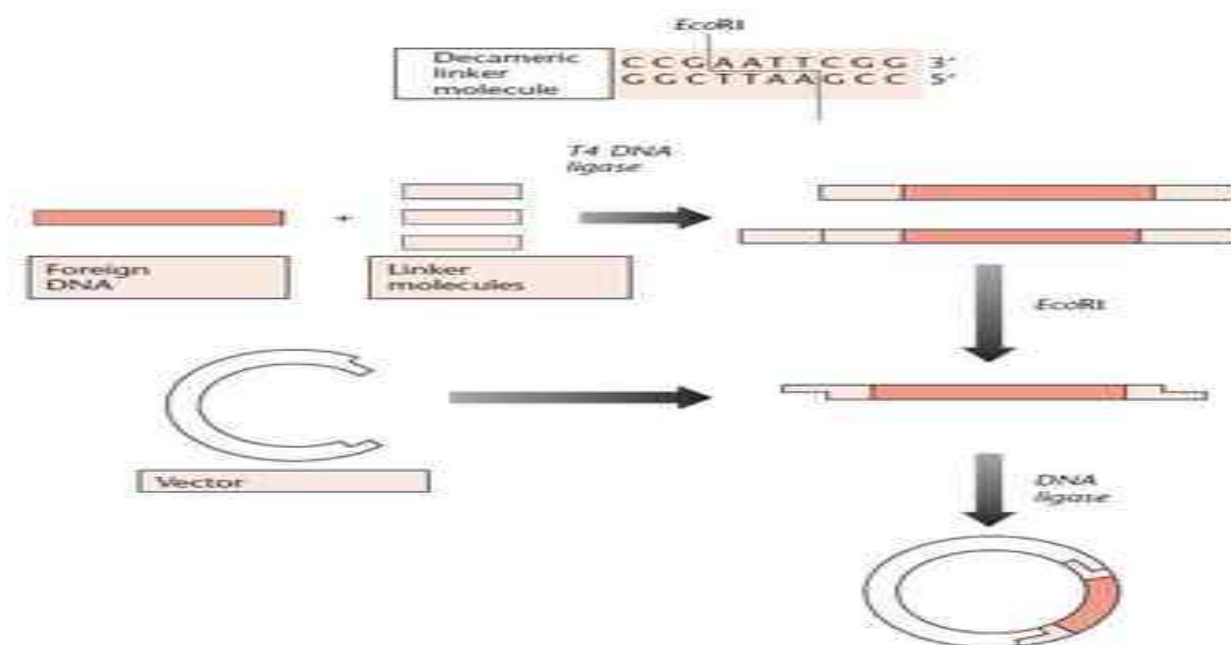




80.81. Blunt end ligation via linker molecules

- 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

Fig: Blunt end ligation by T4 DNA ligase



82. Adaptors

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

- One end of Adaptors is blunt while other end is sticky or cohesive
- At cohesive end, 5' phosphate has been removed and it contains OH group. The removal of phosphate from cohesive end has advantage. That advantage is; when phosphate is not at the cohesive end the adaptor molecules present in the mixture cannot self ligate

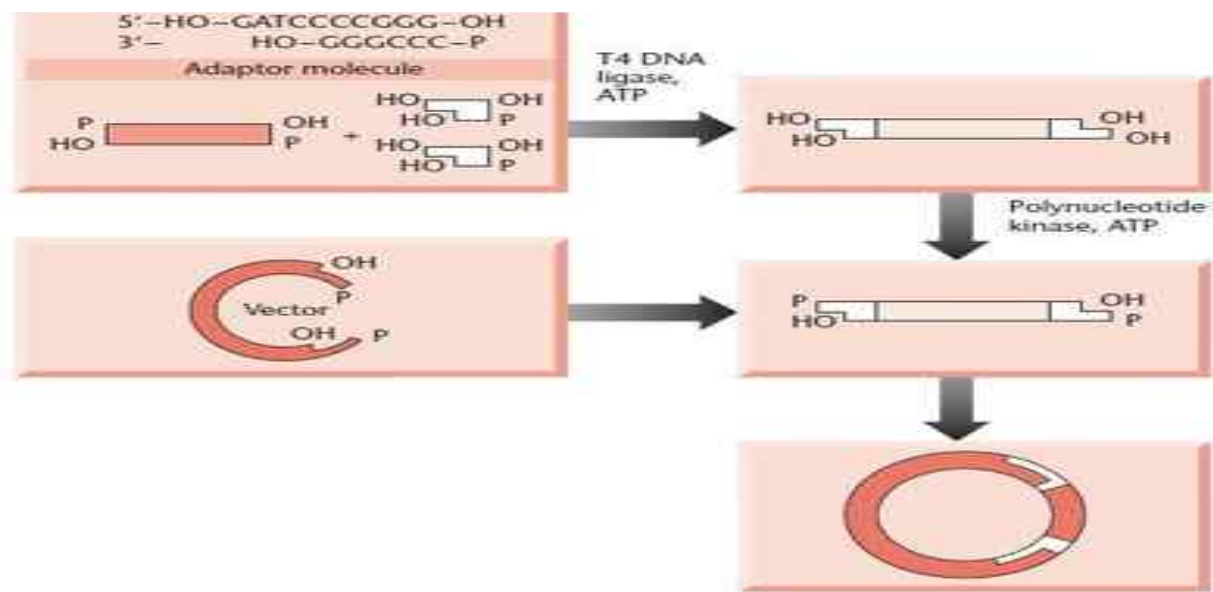


Fig: Use of

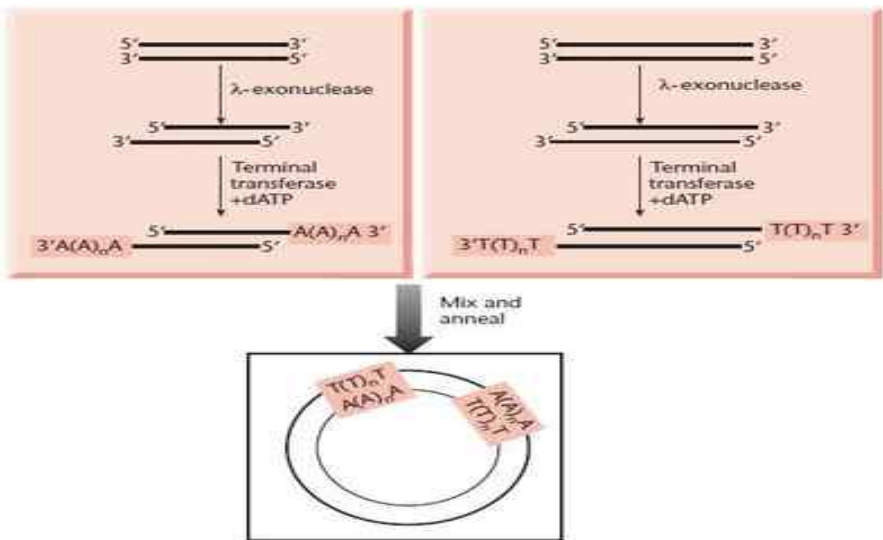
BamH1 adaptor molecule

Difference between linker and adaptor

- 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



Use of calf-thymus terminal transferase to add homopolymer tails to two DNA molecules

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

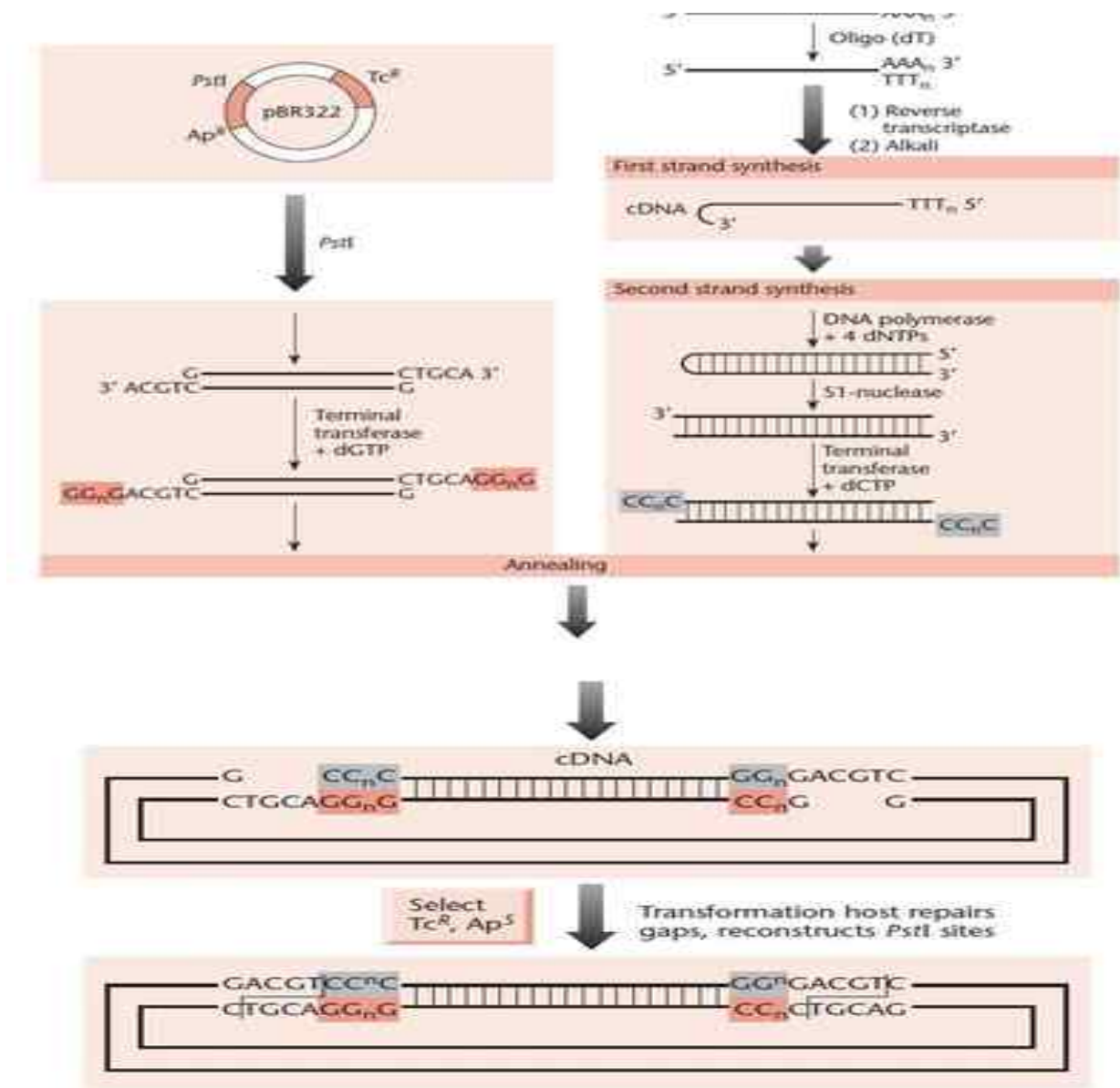


Fig: Insertion of cDNA into vector by homopolymer tailing

Conclusion:

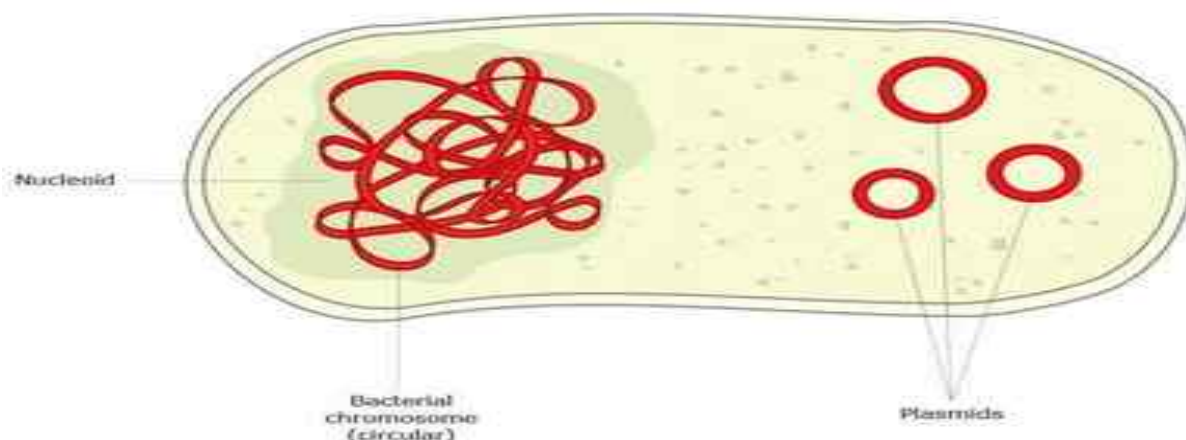
- cutting and joining DNA molecules is basic step in gene manipulation

Basic Biology of plasmid and phage vectors

85.Plasmids

Scientists have developed different type of vector system with the passage of time. Bacterial plasmids are well famous. They can deliver target DNA of the size fifty thousand base pairs or fifty kilo bases. After plasmids lambda phages have also been used to deliver target DNA and they relatively transfer larger DNA fragments as compared to plasmids. Size of target DNA may be 24 thousand bases or 24 kilo bases. Some plasmids that contain cohesive side of the phage vectors have also been developed. These plasmids contain cohesive sides that are necessary for the phage particles packaging. They are called as the Cosmids. They can transfer target DNA having the size of 34 kilo bases to 47 kilo bases. Bacterial artificial chromosomes can deliver larger DNA fragments. To clone large DNA fragments in yeasts, Yeast artificial chromosomes have also been developed. Here we will discuss 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



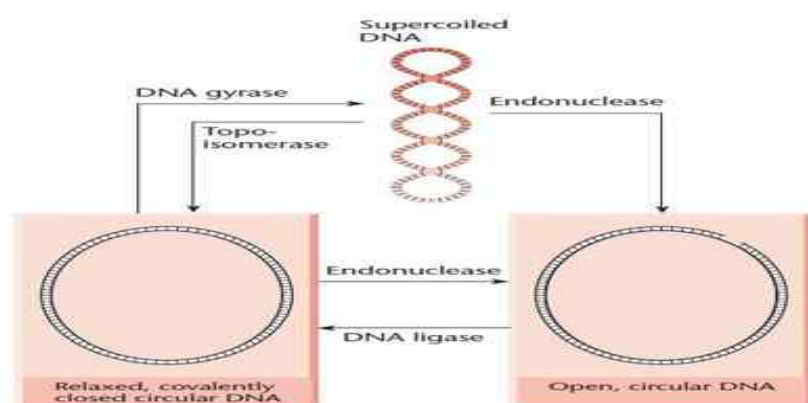
- Plasmid help bacteria to survive stress

Plasmid Biology

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

When we try to isolate plasmid from the bacterial cells, it is present mostly in super coiled form. There are different types of enzymes that can inter convert super coiled DNA into covalently closed circles DNA (CCCDNA) or open circle DNA (OCDNA). This inter conversion of super coiled DNA into CCCDNA or OCDNA mediated by different types of enzymes. For example Topoisomerases can unwind the DNA when DNA replication is going on. So in the direction of movement of replication fork, stress is created on DNA. If this stress is not relieved then DNA replication is affected. By creating artificial leaks topoisomerases can release topological stresses on the DNA. So topoisomerase can also unwind the DNA temporarily. In addition to that if the DNA it is again converted into CCCDNA and if we want to convert it into super coil form we can use DNA gyrase. So it can again wind but in opposite direction to produce negative coiling. If we want to create OCDNA then the DNA can be treated with Endonucleases. So that it can cut one strand of DNA under some situations so that we can have open circle form of DNA.

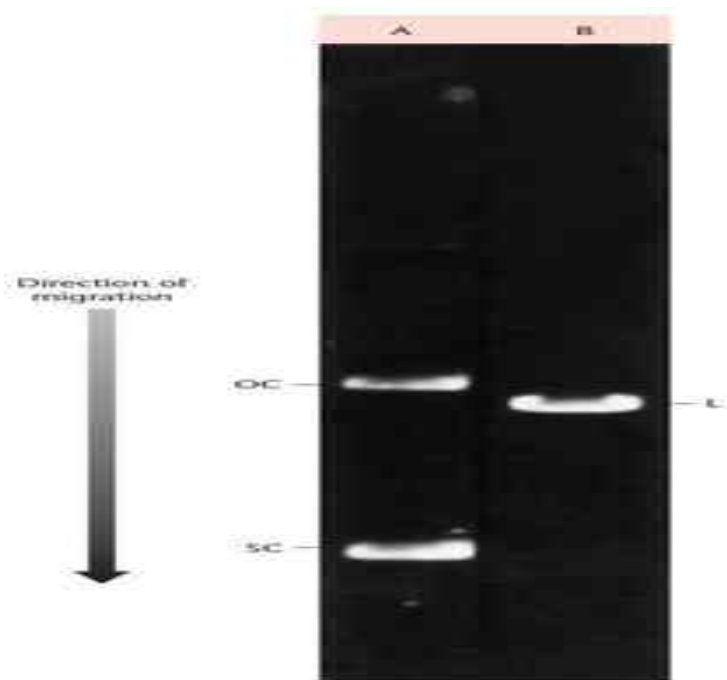


Interconversion of supercoiled, relaxed CCC DNA and OC DNA

Some chemical agents can also influence the super coiling of DNA depending upon its concentration that is used. One of the chemical agents is Ethidium bromide.

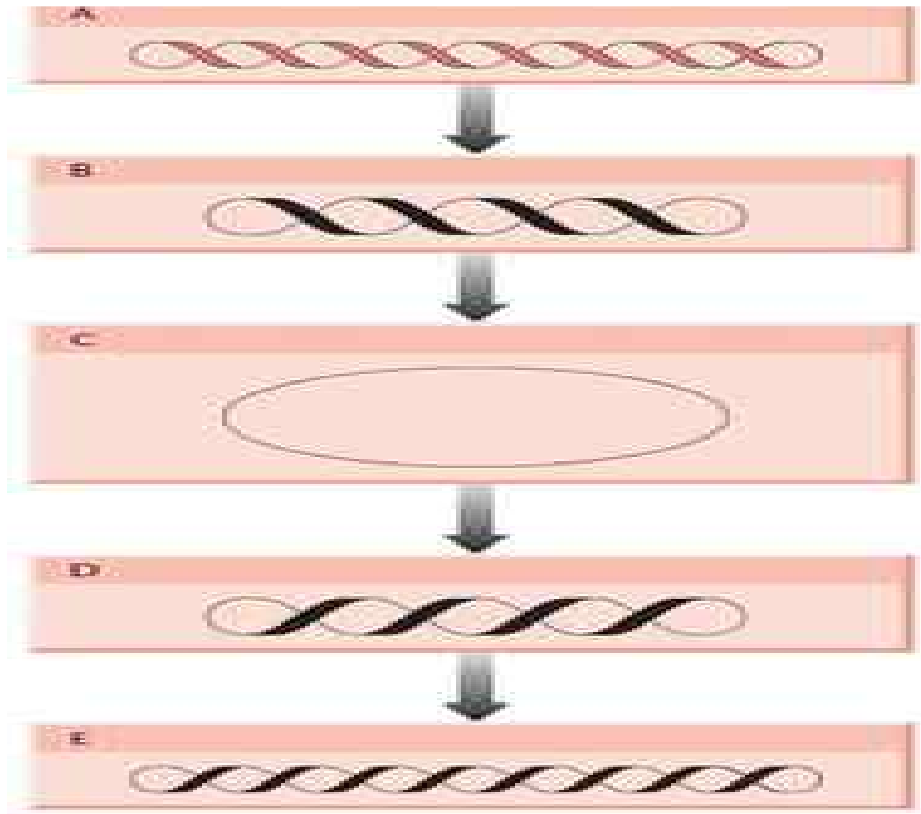
- DNA bands can be visualized by soaking the gel in a solution of ethidium bromide
- As the amount of intercalating agent i.e. ethidium bromide increases, the double helix untwists until the open form of circular DNA is produced

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



Effect of ethidium bromide on supercoiling of DNA

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

88. 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

When plasmids are present, they can exhibit different type of beneficial traits that help microbes to survive in the environment. For example some plasmid contains genes that mediate antibiotic resistance of microbes. Some phenotypic traits exhibited by plasmid-carried genes;

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

cryptic plasmids:
Plasmids to which phenotypic traits have not been ascribed are called **cryptic plasmids**.

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
- i.** Conjugative plasmid
- ii.** Non-conjugative plasmid

Conjugative plasmids are those that can be transfer from one bacterium to another bacterium by a process which is called as conjugation. This process is mediated by those plasmids that contain a region of the DNA which is called as *tra* region that contain gene that mediate transfer of plasmid DNA from donor to recipient. Those plasmids that do not contain *tra* genes they are called as **non-conjugate plasmids**. Another classification is based upon copy number of plasmids that how many copies of the plasmids are present in a particular bacteria or microorganism. Plasmids may be of high copy number or they can also be called as **relax plasmids** (Copy number high). If copy number is low then plasmids are called as **stringent plasmids**. Those plasmids that mediate conjugation, they are high molecular weight and they are present in low copy number.

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

Plasmid	Size (Mda)	Conjugative	No. of plasmid copies	Phenotype
ColE1	4.2	No	10-15	ColE1 production
RSF1030	5.6	No	20-40	Ampicillin resistance

cloDF13	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 307	65	Yes	1-3	Enterotoxin production

90.Host range of plasmids

- When plasmids are present in microbes, whether they have narrow host range or wide host range.
- Plasmids encode only a few of the proteins required for their own replication
- Host range of plasmid is determined by *ori* region (Point where DNA replication mainly starts)
- Plasmids whose *ori* derived from Col E1 have restricted host range
- 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.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*

Properties of some ‘natural’ plasmids used for cloning DNA

- Cloning in pSC101, ColiE1 and RSF2124 (There are certain limitations in using these plasmids as cloning vehicles).

Plasmid	Size (Mda)	Sites of endonucleases	Selectable markers	Insertional inactivation
pSC101	5.8	<i>Xho</i> I, <i>Eco</i> RI, <i>Pvu</i> II, <i>Hinc</i> II, <i>Hpa</i> I	Tetracycline resistance	---
		<i>Hind</i> III, <i>Bam</i> HI, <i>Sal</i> I	—	Tetracycline resistance
Col E1	4.2	<i>Eco</i> RI	Immunity to colicin E1	Colicin E1 production
RSF2124	7.4	<i>Eco</i> RI, <i>Bam</i> HI	Ampicillin resistance	Colicin E1 production

94.Maximizing the expression of cloned gene

Structural instability

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

95.Desirable properties of plasmid cloning vehicles

- An ideal cloning vehicle would have the following three properties

phenotypic traits on host cells **iii)** Single sites for number of restriction endonucleases

Lecture 96

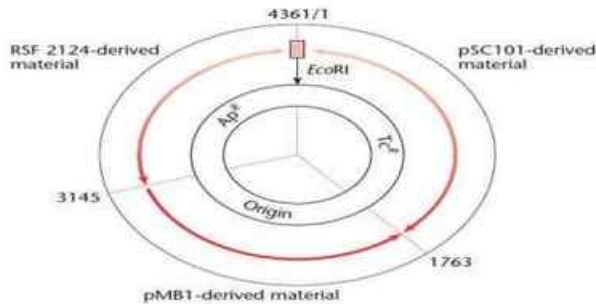
Plasmid Biology:
Use of pSC101 for cloning

- Expression of *Staphylococcus* plasmid genes in *E. coli*
- pSC101 contain Tetracycline resistance *P1248* from *Staphylococcus*

Lecture 97

pBR322, a purpose-built cloning vehicle:

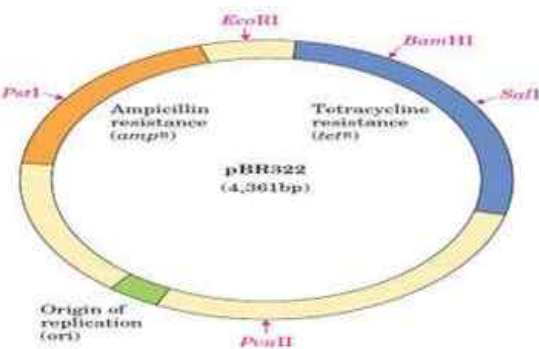
- pBR322 is an example of *in vitro* constructed cloning vehicle
- pBR322 contains the *Ap^R* and *Tc^R* genes of RSF2124 & pSC101, combined with replication element of pMB1



The origin of plasmid pBR322

pBR322, a purpose-built cloning vehicle

- It has 40 target sites for different restriction enzymes

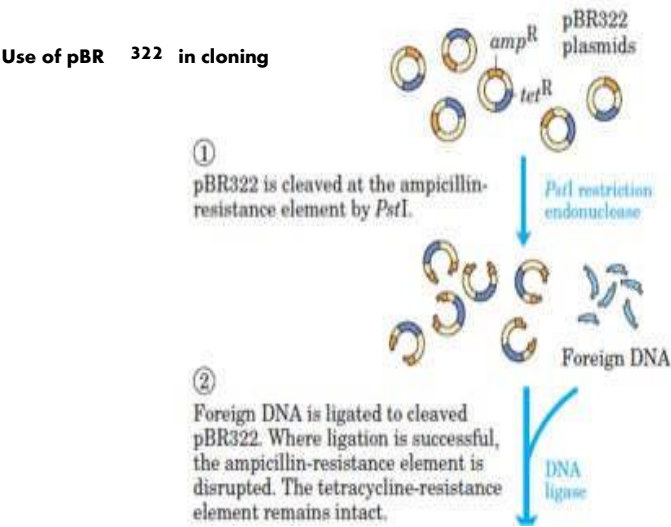


The constructed *E. coli* plasmid pBR322 pBR322, a purpose-built cloning vehicle □ It is difficult to clone DNA segments longer than about 15,000 bp when plasmids are used as the vector

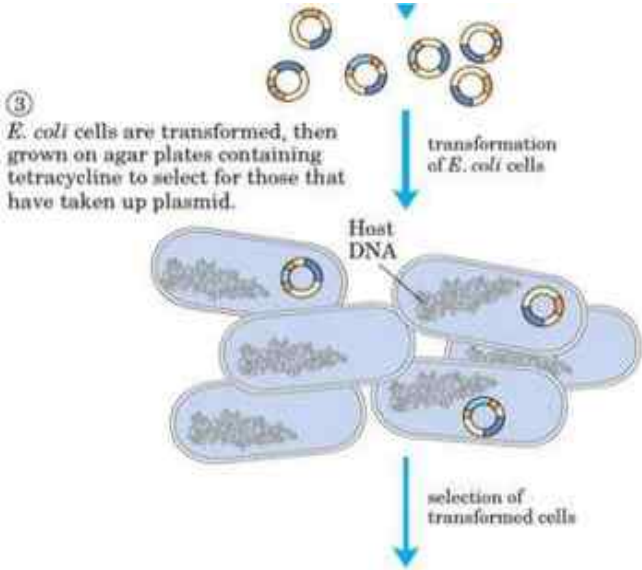
Lecture 98

Cloning with pBR322

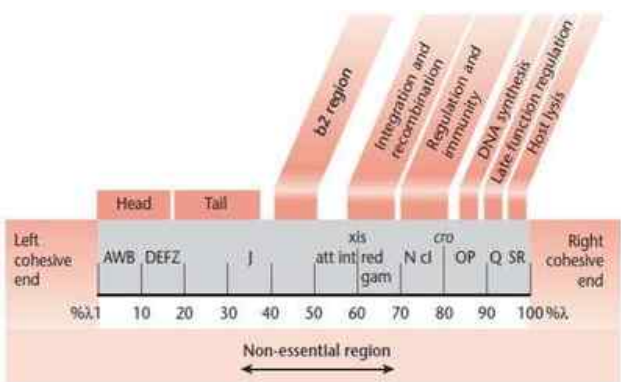
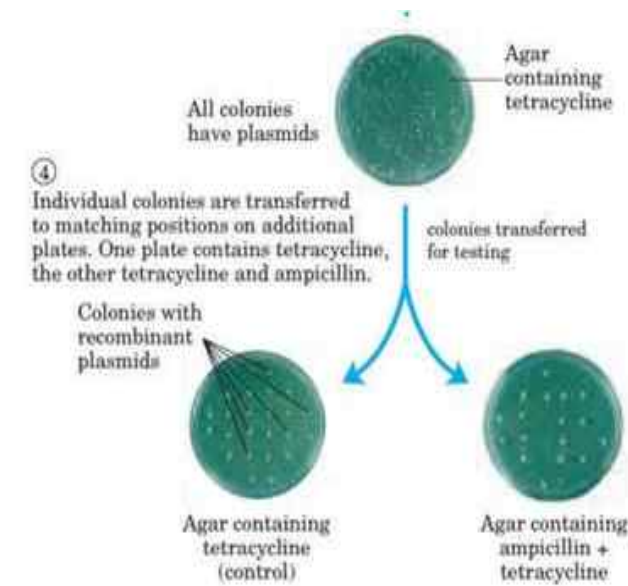
- Plasmid pBR322 has a number of characteristics that make it suitable as cloning vehicle.



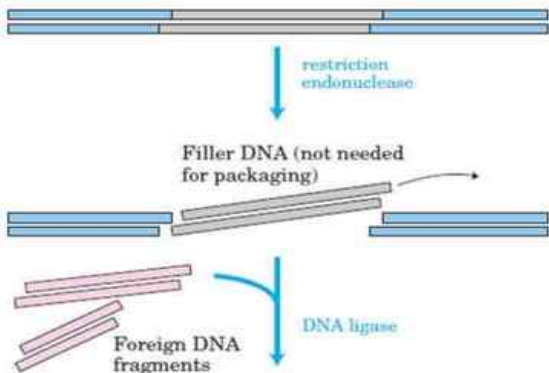
Use of pBR322 in cloning



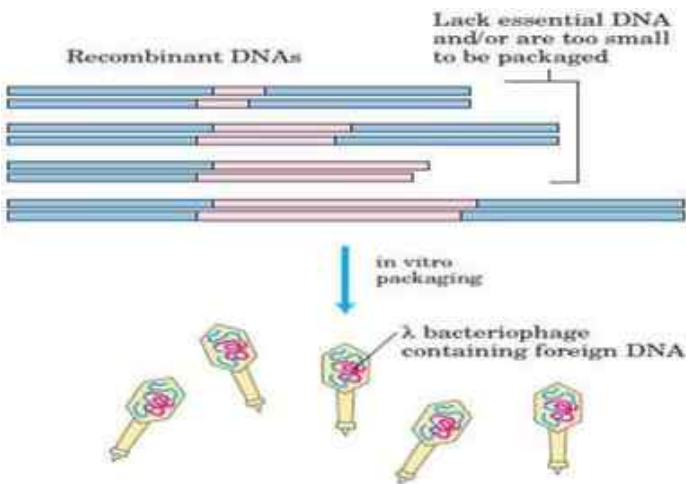
Use of pBR322 in cloning



Map of λ chromosome



Bacteriophage λ cloning vectors



Bacteriophage λ cloning vectors

Lecture 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

Runaway plasmid vectors

- Loss of copy number control –runaway replication

Lecture 100

Runaway plasmid vectors

- Loss of copy number control – runaway replication

Lecture 101

Phage vectors 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

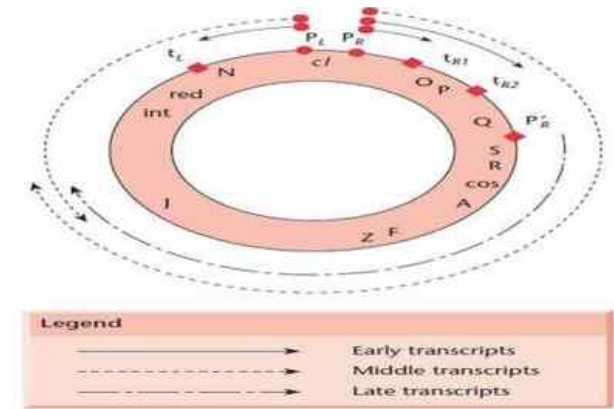
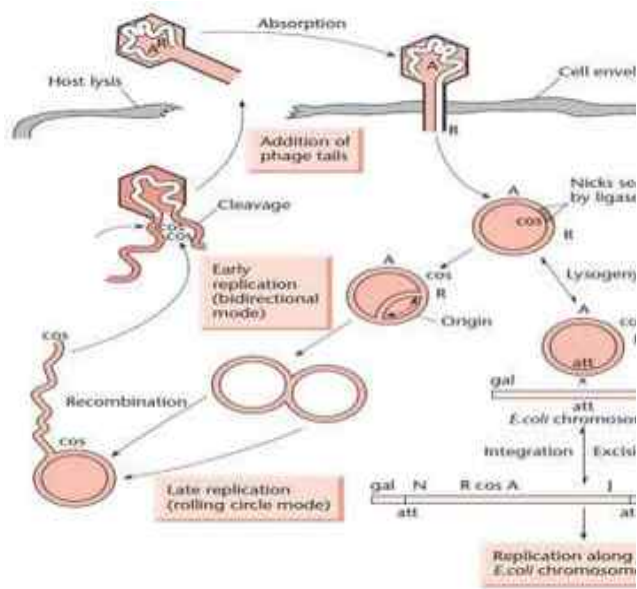
Lecture 102

Replication of phage-λ DNA in lytic & 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



Major promoters & transcription termination sites of λ

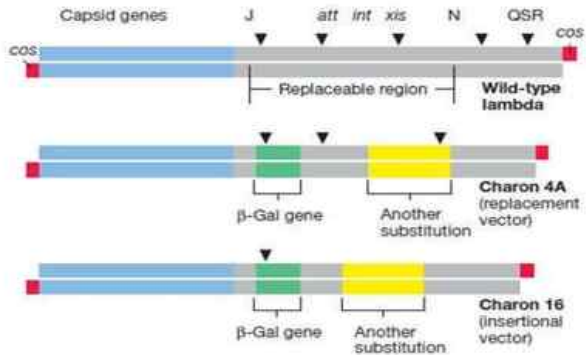
Lecture 103

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 & 16 are both derivatives of lambda with various substitutions & deletions in the nonessential region

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



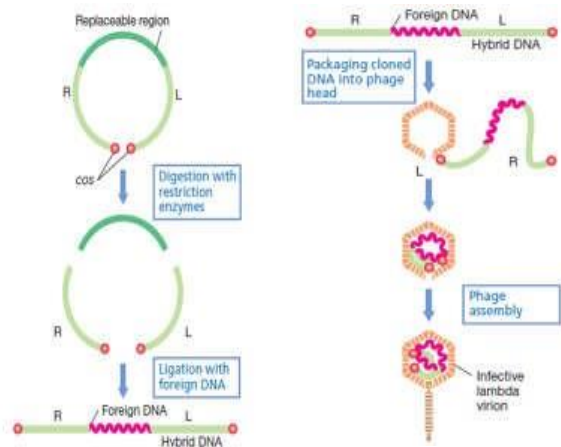
Modified λ cloning vectors

Lecture 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

Bacteriophage lambda as cloning vector



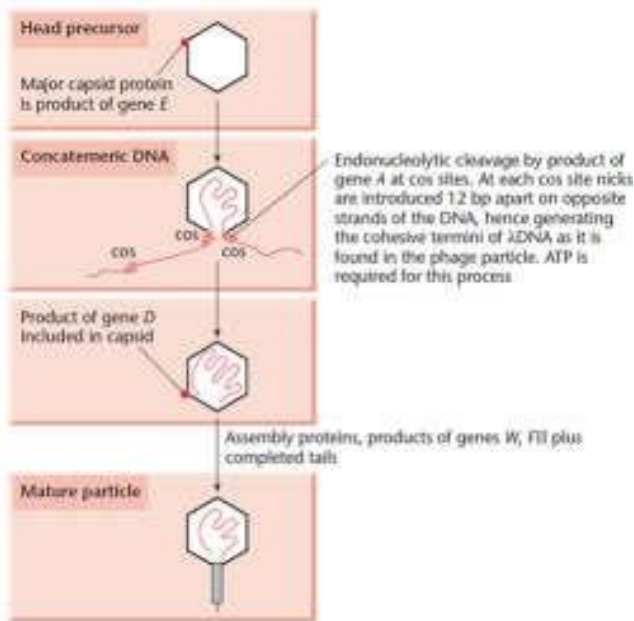
- Both Charon vectors are also engineered to contain reporter genes such as for β -galactosidase for detection of recombinant phages.

Lecture 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/ μ g of vector DNA

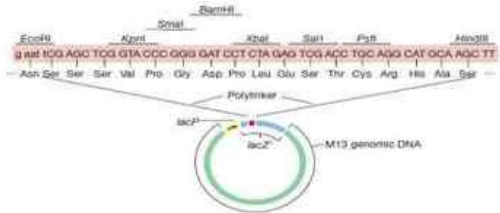
Packaging of phage- λ DNA into phage particles



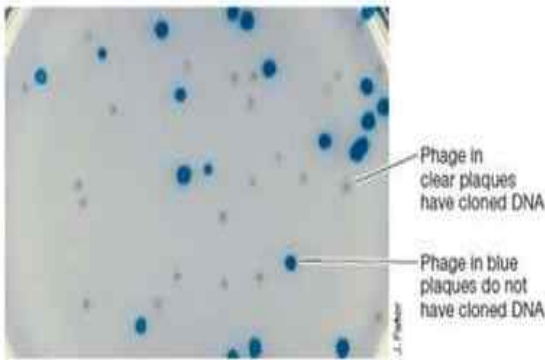
Lecture 106

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



Cloning using bacteriophage M13mp18



Vectors for DNA sequencing: bacteriophage M13

- Use of M13 in molecular cloning

Lecture 107

Cosmid vectors:

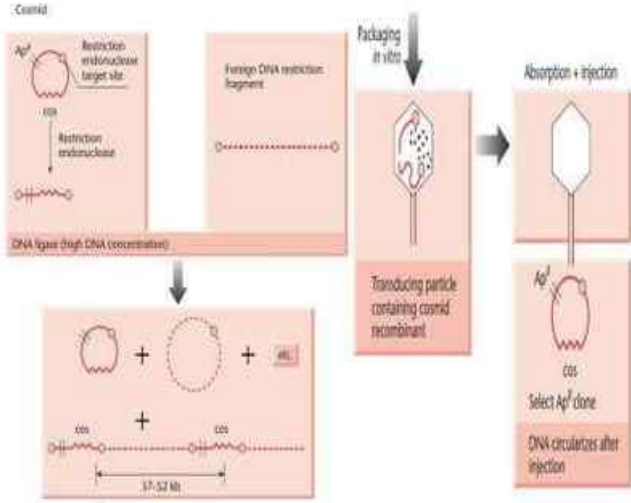
- Plasmids have been constructed which contain a fragment of λ DNA including 'cos' site
- These plasmids have been termed cosmids and can be used as gene-cloning vectors in conjunction with the *in vitro* packaging system.

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

- Cosmids are particularly attractive vectors for constructing

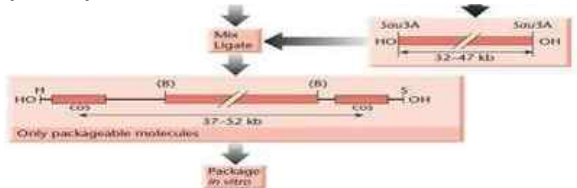
libraries of eukaryotic genomic fragments.

Simple scheme for cloning in a cosmid vector

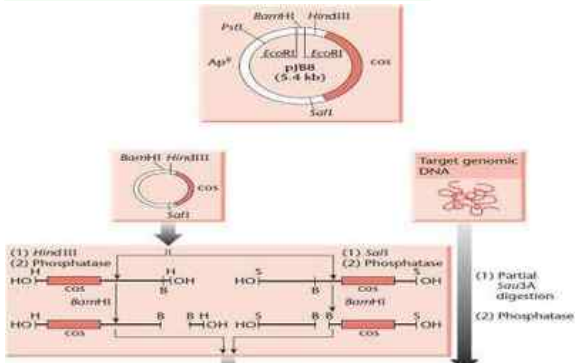


Lecture 108

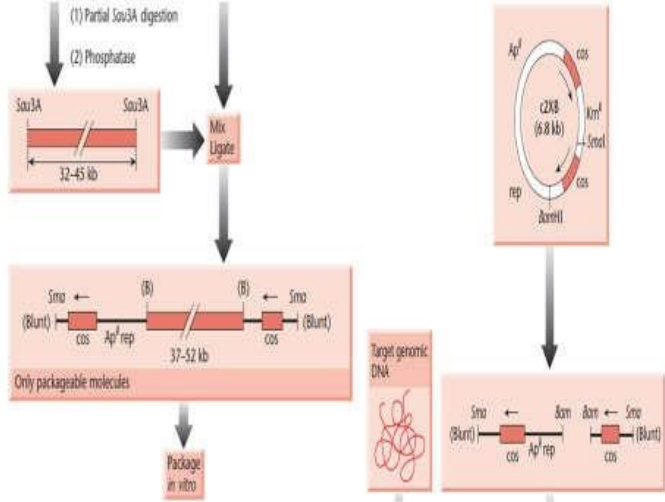
Modified schemes for cloning in Cosmid vectors □ Difficulties in cloning with cosmids have been overcome in a cosmid cloning procedure devised by IshHorowicz and Burke (1981). An alternate solution to these problems has been devised by Bates and Swift (1983) Who constructed cosmid c2XB



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



Cosmid cloning scheme of Bates and Swift (1983)

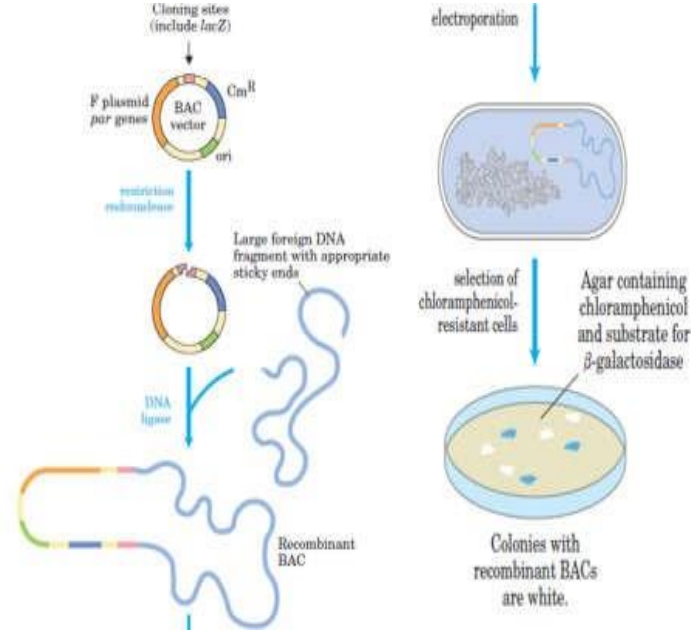


Lecture 109

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

Lecture 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

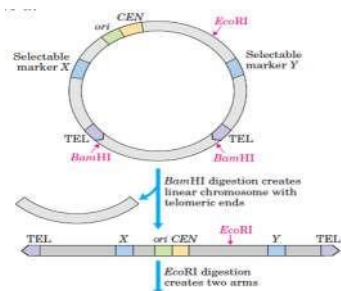


Lecture 111

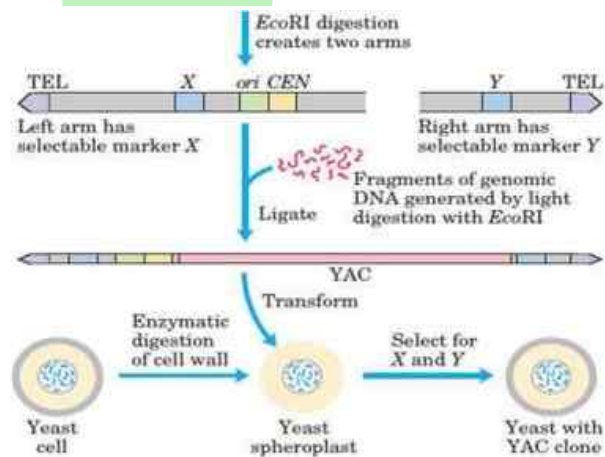
Yeast artificial chromosomes (YACs):

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

YACs with inserts more than 150,000 bp are nearly as stable as normal cellular chromosomes



Construction of yeast artificial chromosome

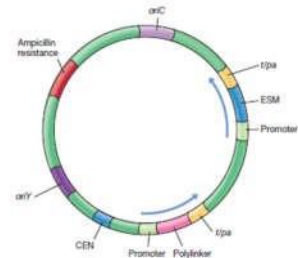


Lecture 112

Shuttle and Expression vectors:

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

To control the expression of cloned genes expression vectors are used.



Genetic map of a shuttle vector used in yeast.

Lecture 113

Comparison of different cloning vectors-summary

Different cloning vectors can carry variable sizes of target DNA into host organisms.

Table. Maximum DNA insert possible with different cloning vectors

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>	$\leq$ 300 kb
YACs	<i>S. cerevisiae</i>	200-2000 kb

Lecture 114

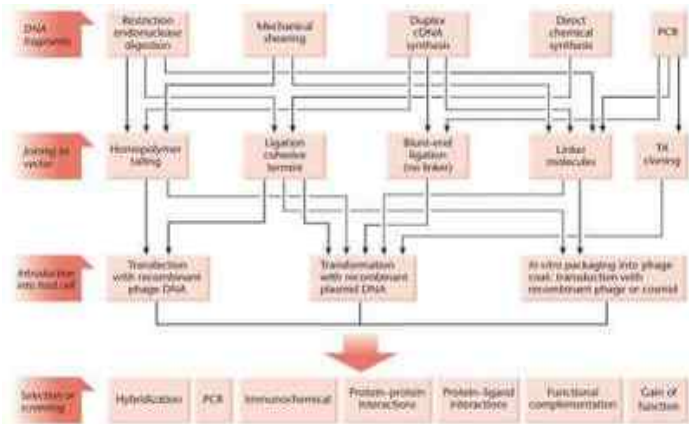
OVERVIEW OF CLONING

STRATEGIES:

- Any cell based cloning procedure has 4 essential parts i). A method of generating DNA fragment for cloning ii). A reaction that inserts that fragment into the chosen cloning vectors. iii). A means for introducing that recombinant vector into a host cell wherein it is replicated iv). A method of selecting recipient cells that have acquired the recombinant

Introduction:

- There are two major strategies for isolating sequences from complex sources i). Cell based cloning strategy, is to divide the source DNA into manageable fragments and clone everything ii). The second strategy is to selectively amplify the target sequence directly from the source DNA by using PCR



Generalized overview of cloning strategies

Lecture 115

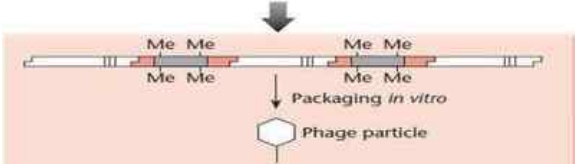
Cloning genomic DNA:

Genomic DNA libraries:

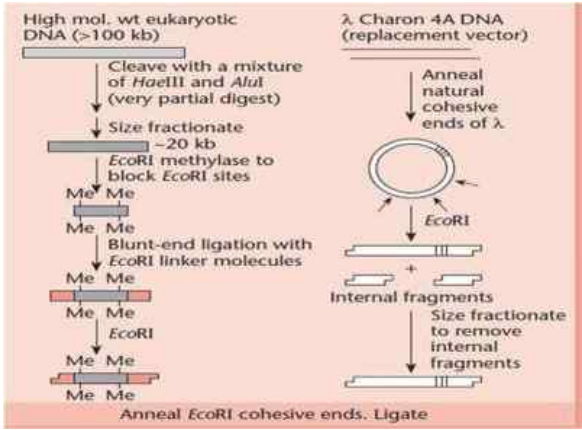
- Producing representative genomic libraries in λ cloning vector. Table 1. Genome size of selected organisms

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>	1.5×10^9
<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

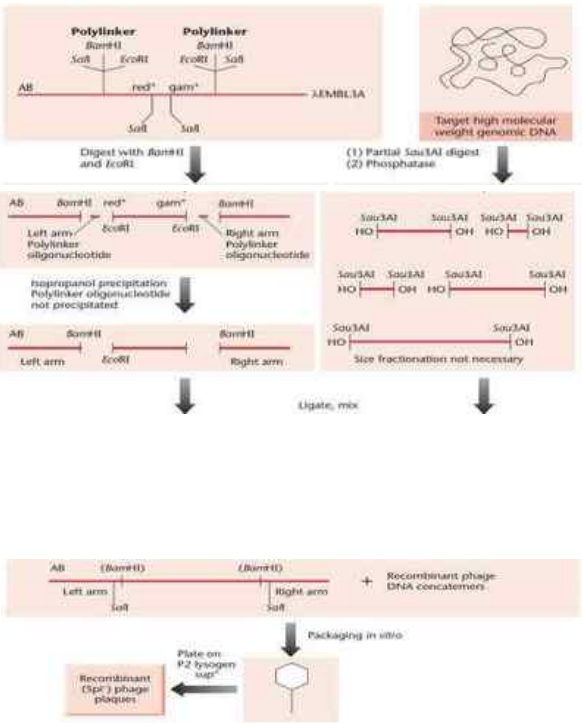
- Cloning of large DNA fragments of 20 kb in λ replacement vectors, fewer clones are required for nearly complete library.
- How can appropriately sized random fragments be produced?
 - Average fragment size can be controlled by mechanical shearing
 - The more commonly used procedure is restriction endonucleases



Maniatis' strategy for producing a representative gene library.



Creation of genomic DNA library using the phage-λ vector EMBL3A



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.

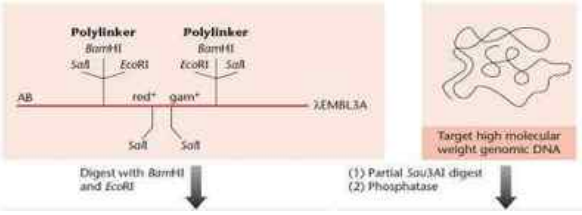
PCR as an alternative to genomic DNA cloning

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

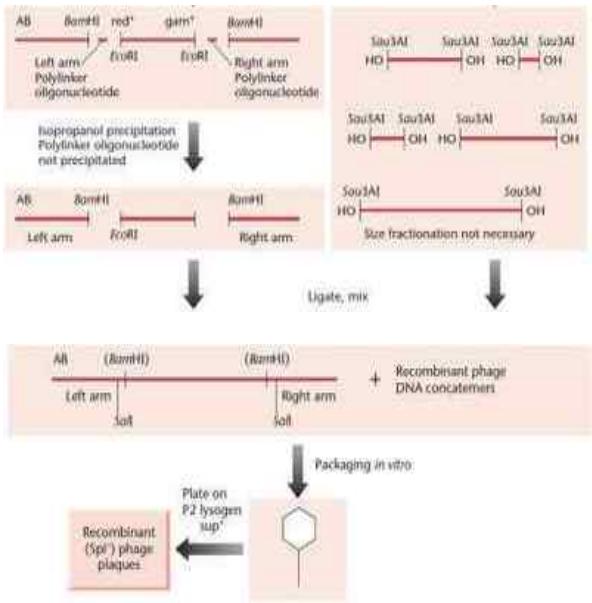
Lecture 116

λEMBL vectors for genomic library construction

High molecular weight genomic DNA is digested with Sau3AI and subsequently ligated in λEMBL vector digested with BamHI



Creation of genomic DNA library using the phage-λ vector EMBL3A



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

Phage-λ vectors for cDNA cloning and expression

• λgt10 and λgt11 vectors

Table. Abundance classes of typical mRNA populations

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
	7	4 000
	12 500	5

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

Lecture 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.

Lecture 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.

Lecture 119

cDNA cloning

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

Lecture 120

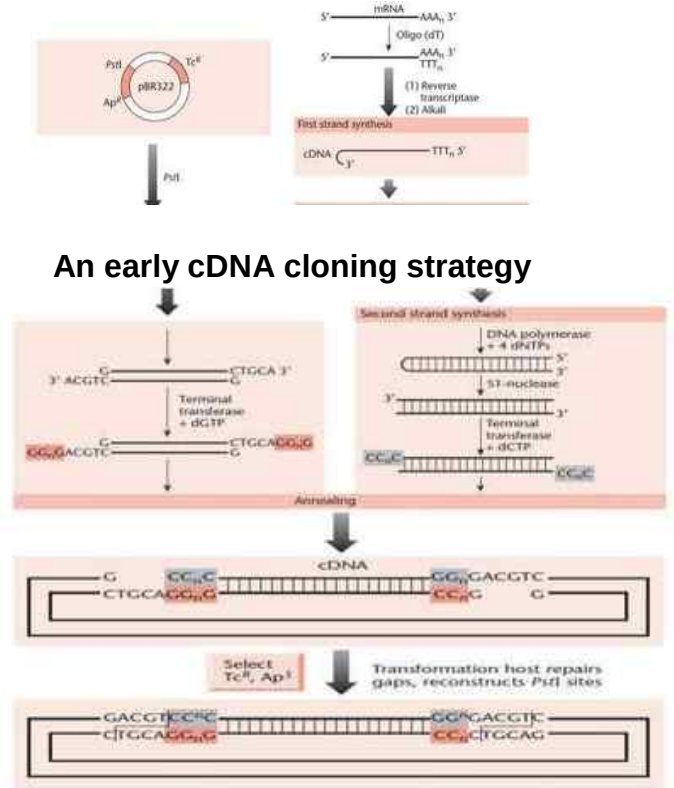
cDNA libraries

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

Lecture 121

Preparation of cDNA for library construction:

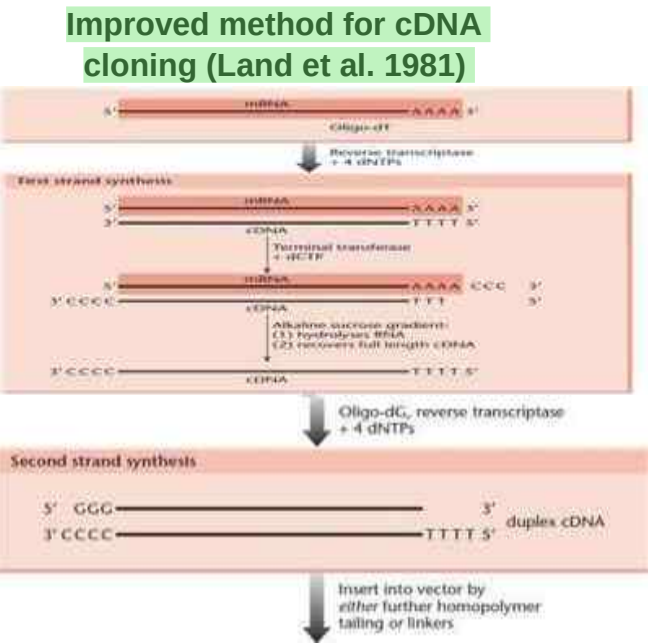
- The cDNA synthesis reaction
- Development of cDNA cloning strategies



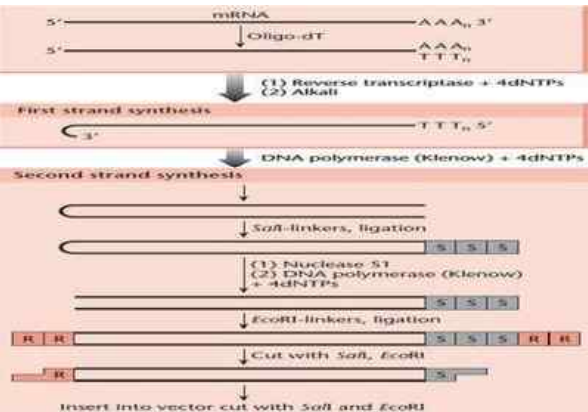
Lecture 122

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



Method for direction cDNA cloning



Lecture 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.

Lecture 124

Screening strategies:

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

Selection for the presence of vector:

- 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.

Lecture 125

Screening by hybridization

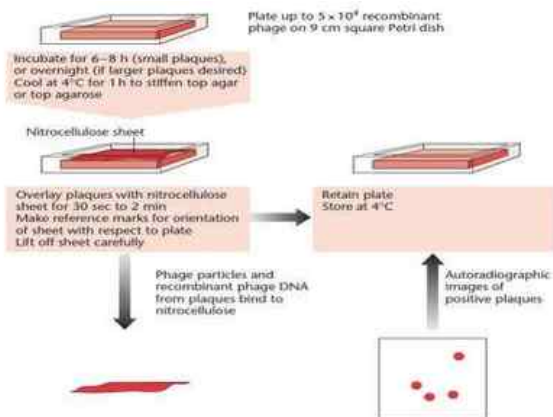
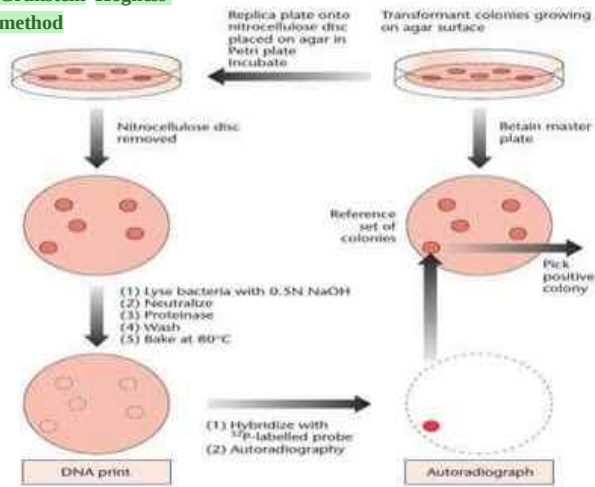
- Nucleic acid hybridization is the most commonly used method of library screening.
- Grunstein and Hogness (1975) developed a screening procedure to detect DNA sequences in transformed colonies by *in situ* hybridization.
- The results of the hybridization can be monitored by autoradiography

Lecture 126

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

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

Grunstein-Hogness method



Lecture 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.

Lecture 128

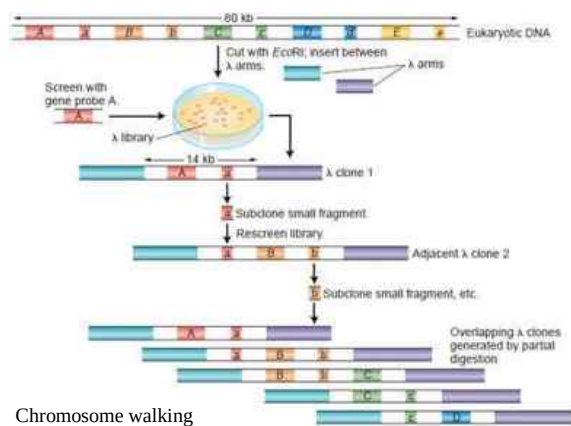
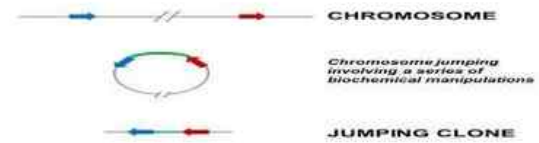
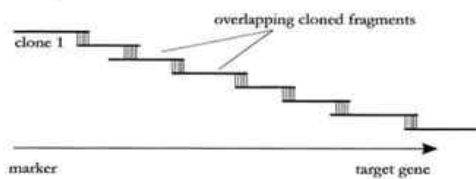
Chromosome walking:

- 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.

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
- One drawback to this method is the requirement that each DNA segment used is not repeated elsewhere in the genome

Chromosome walking



Chromosome walking

Lecture 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

Lecture 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 genespecific 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

Lecture 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

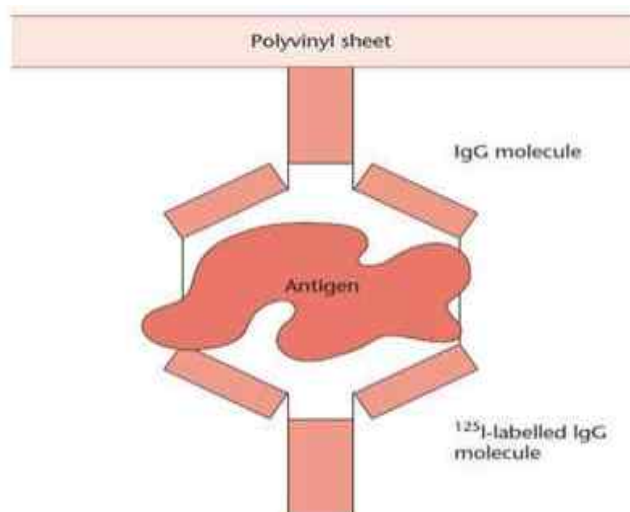
Lecture 132

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 ^{125}I by iodination *in vitro*

Immunochemical detection method of Broome and Gilbert



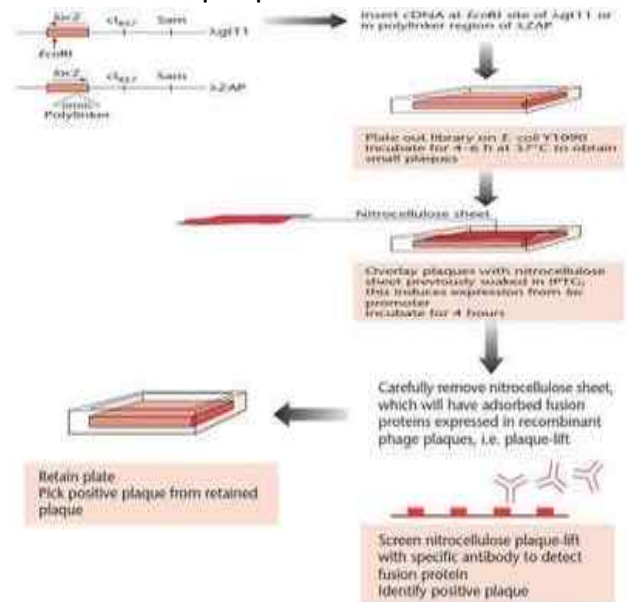
Lecture 133

Immunochemical screening of $\lambda\text{gt}11$:

- 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 nonspecific signal

25 December 2017

Immunochemical screening of $\lambda\text{gt}11$ or λZAP recombinant plaques



Lecture 134

South-western and north-western blotting

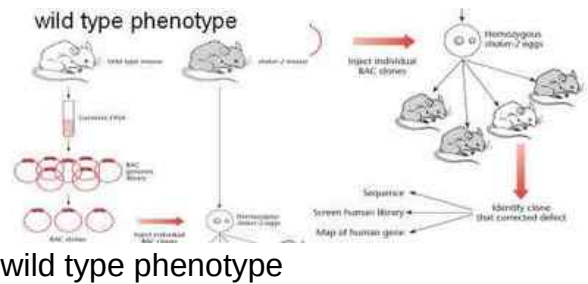
- It involves the screening & isolation of clones expressing sequence specific DNA/RNA-binding proteins
- Screening is carried out without using an antibody, by incubating the membranes with radiolabelled double stranded DNA probe

Lecture 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.
- Functional complementation is the process by which a particular DNA sequence compensates for a missing function in a mutant cell and thus restores the



Lecture 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

Lecture 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

Lecture 138

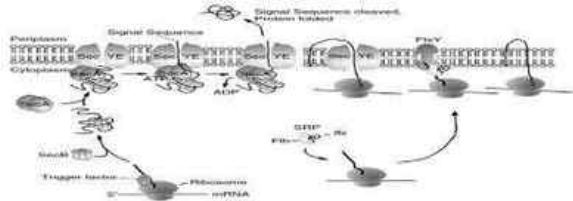
Expression in *E. coli* of cloned DNA

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

Figure. General secretion pathway

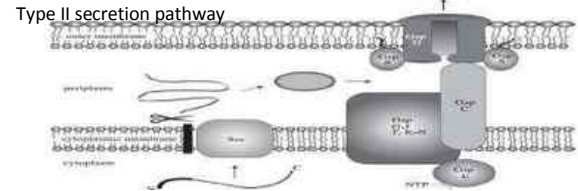
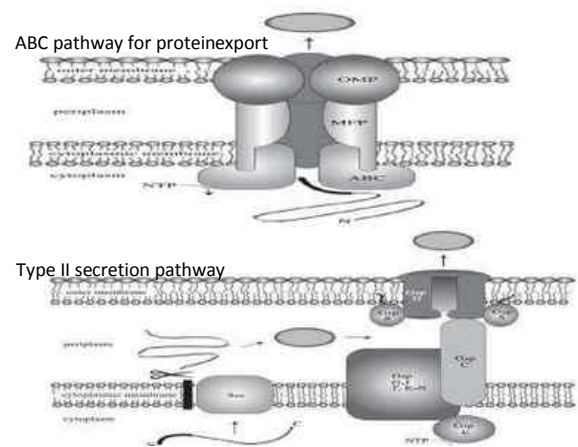


- There are actually several general mechanisms of protein export in bacteria
- Each system transport a particular type of proteins

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 **Constructing the optimal promoter**

- Large No. 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-) & -10 region or Pribnow box (5'-TATAAT-)



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

	-35 Region		-10 Region
	1 2 3 4 5 6 7 8 9	10 11 12 13 14 15 16 17	
CONSENSUS	•••TTGACA••••••••••	••••••••••TATAAT••	
lac	GGCTTTACACTTATGCTTCGGCTCGTATATTGT		
trp	CTGTTGACAATTAATCAT	CGAACTAGTTAACTAG	
λP_L	GTGTTGACATAAATACCA	CTGGCGGTGATACTGA	
tecA	CACCTGATACTGTATGAA	GCATACAGTATAATTG	
tacI	CTGTTGACAATTAATCAT	CGGCTCGTATAATGT	
tacII	CTGTTGACAATTAATCAT	CGAACTAGTTTAAATGT	

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

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.

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

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

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

Host cell physiology can affect the level of expression

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

Lecture 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.

Lecture 140**Constructing the optimal promoter**

- Large No. 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-) & -10 region or Pribnow box (5'-TATAAT-)
- Expression from a strong promoter can represent

20-40% of cloned gene product of total cell protein x

Lecture 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

Lecture 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

Lecture 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

Lecture 144**Plasmid stability**

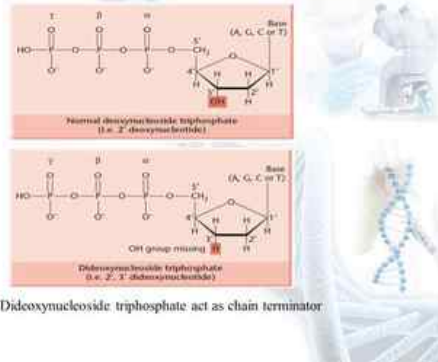
- Having maximized the expression of a particular gene it is important to consider what effects this will have on the bacterium harbouring the recombinant plasmid □ The loss of plasmids due to defective partitioning is called segregative instability

Lecture 149 Chain termination or dideoxy procedure

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

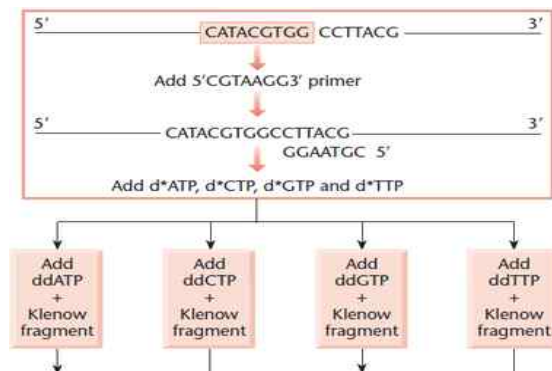
Analysing DNA sequences



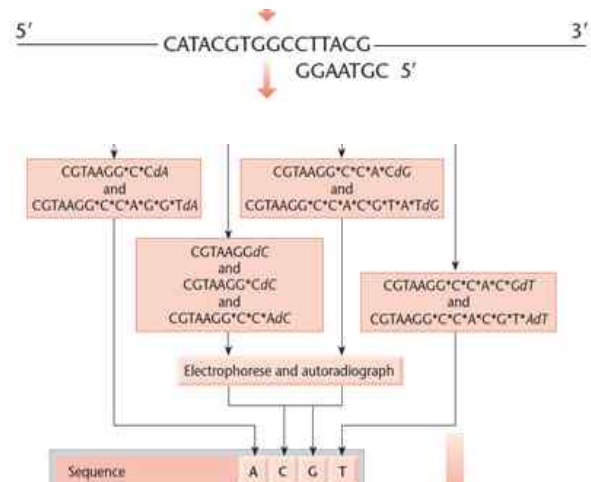
Chain termination or dideoxy procedure

- DNA synthesis is carried out in the presence of the four deoxynucleoside triphosphates, one or more of which is labelled

with ^{32}P and in four separate incubation mixes



DNA sequencing by chain terminator method



Lecture 150

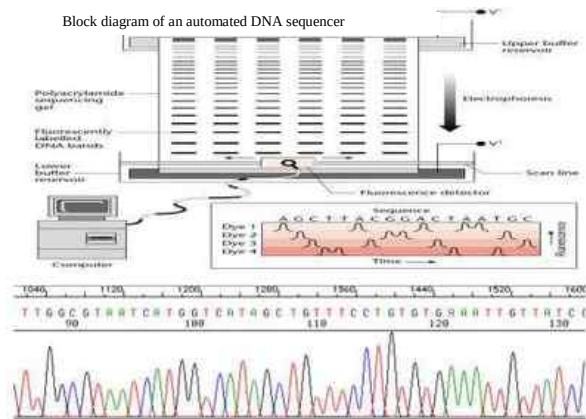
Modifications of chain terminator sequencing:

- 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

Lecture 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 offers a number of advantages



Lecture 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

Lecture 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

Lecture 154

Changing genes: site-directed mutagenesis

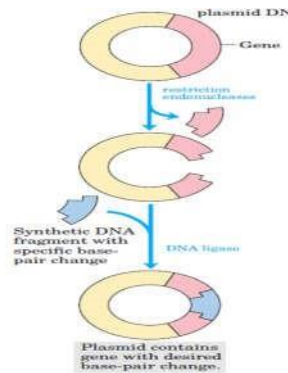
Mutagenesis:

- 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
- Three different methods of site-directed mutagenesis have been devised
 - Cassette mutagenesis
 - Primer extension
 - Procedures based on PCR

Lecture 155

Cassette mutagenesis:

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



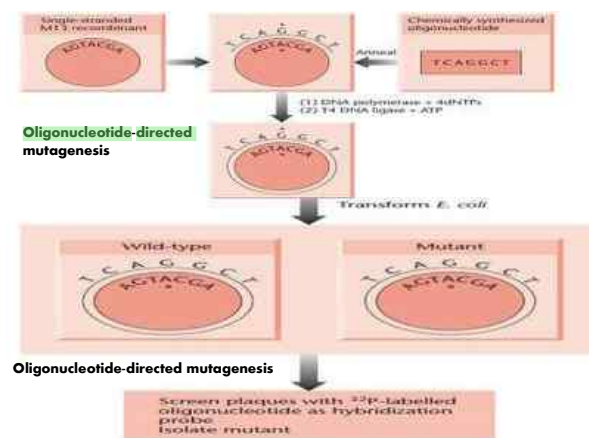
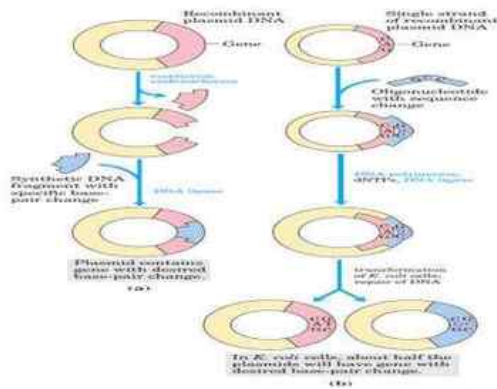
Lecture 156

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.

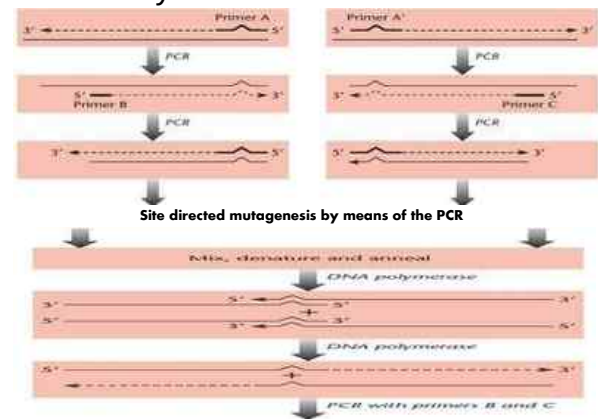
After *E. coli* transformation, heteroduplex that are formed are either that of original wild type DNA or that containing the mutated base.

Two approaches to site-directed mutagenesis



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

■ The advantage of a PCR-based mutagenic protocol is that the desired mutation is obtained with 100% efficiency



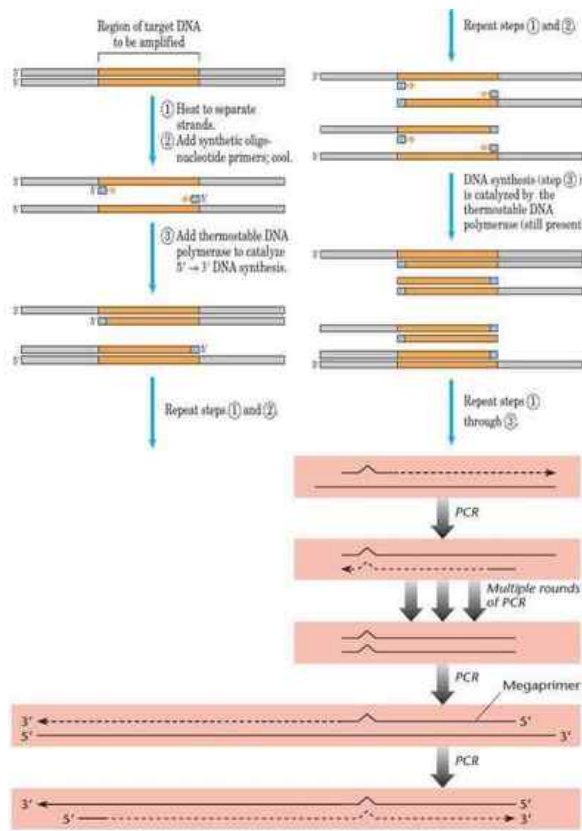
Site directed mutagenesis

by means of the PCR

Lecture 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

The megaprimer method of mutagenesis



Lecture 158

Basic PCR reaction:

- PCR is a technology in molecular biology used to amplify a single or a few copies of a target DNA across several orders of magnitude, generating millions of copies of a particular DNA sequence.
- PCR can also be used to generate DNA fragments for gene cloning by using primers with specific restriction sites

- PCR is a technology in molecular biology used to amplify a single or a few copies of a target DNA across several orders of magnitude, generating millions of copies of a particular DNA sequence.

Lecture 159

PCR principles and procedures:

- PCR amplifies a specific target region of DNA that may range from a few to several kilo base pairs
- Basic reaction require several components
- PCR cycles may consists of several steps of denaturation, annealing and extension.

Lecture 160

DNA polymerases:

- DNA polymerases are enzymes that synthesize DNA molecules by adds nucleotide, in building blocks of DNA

Taq DNA polymerase:

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

Table. Comparison of DNA polymerases

of *E. coli*

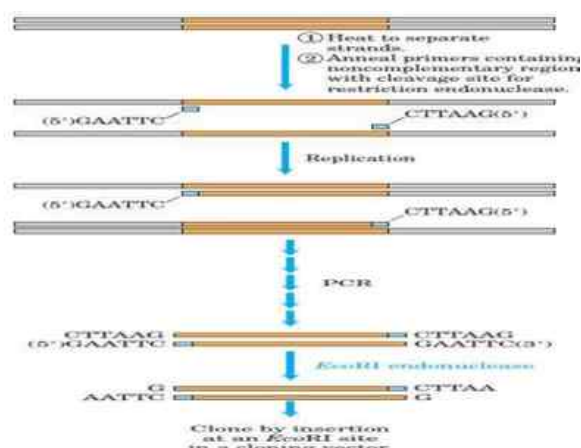


Table. Sources of thermostable DNA

polymerases

Lecture 161

Primers:

□ Primer is a short strand of oligonucleotide that serves as a starting point for DNA synthesis.

Lecture 162

Degenerate primers:

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

Table. Amino acids by codon specificity

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)			

Table. Standard nucleotide coding system

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

Lecture 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

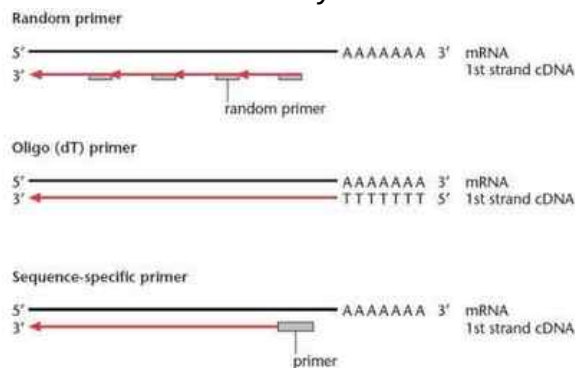


Fig. Strategies for synthesis of first-strand cDNA

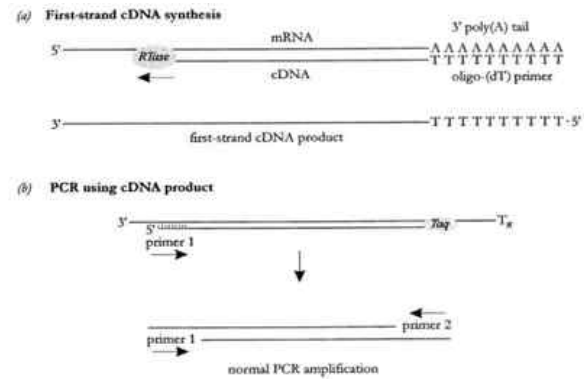
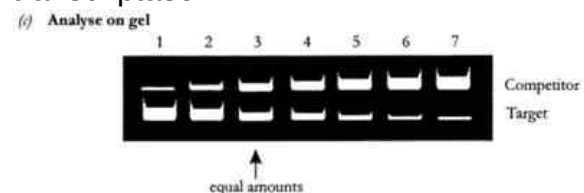


Fig. RT-PCR

Lecture 164

Competitor RT-PCR

- One use of RT-PCR is in determining the amount of mRNA in a sample (competitor RT-PCR) **RT-PCR**
- Reverse transcriptase PCR (RT-PCR) was designed to amplify RNA sequences (mainly mRNA) through synthesis of cDNA by reverse transcriptase



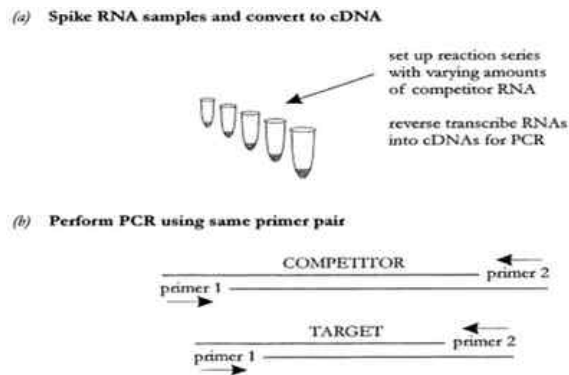


Fig. Competitor RT-PCR

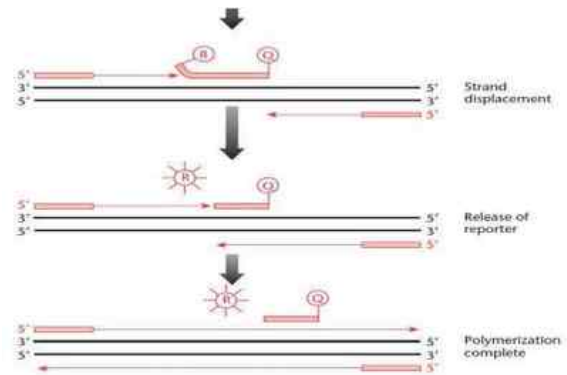
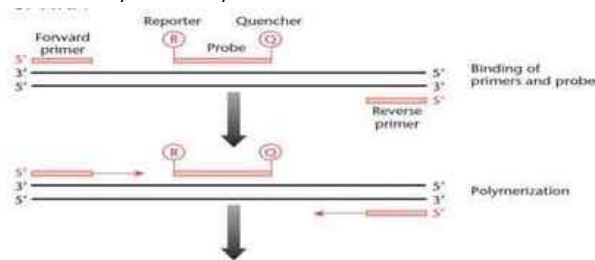


Fig. Real-time quantitative PCR

Lecture 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

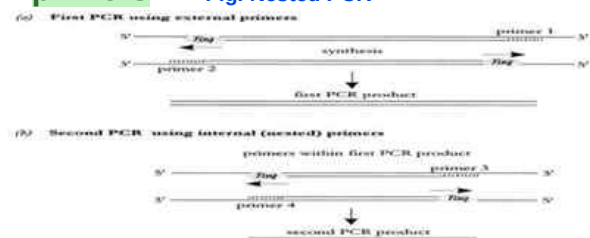


Lecture 166

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**.

Fig. Nested PCR

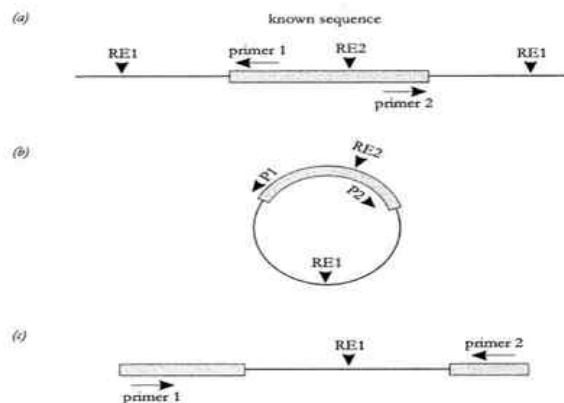


Lecture 167

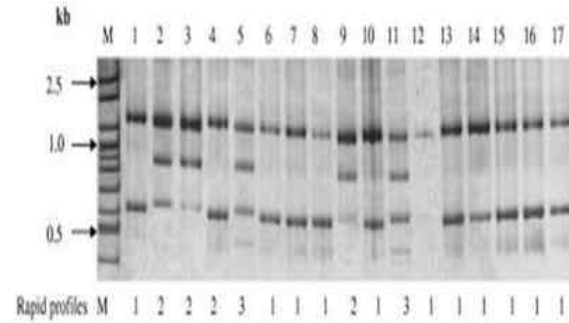
Inverse PCR:

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

Fig. Inverse PCR



reaction but the segments of DNA that are amplified are random.



Lecture 168

Multiplex PCR

□ Multiplex PCR is used to amplify several different DNA sequences simultaneously.

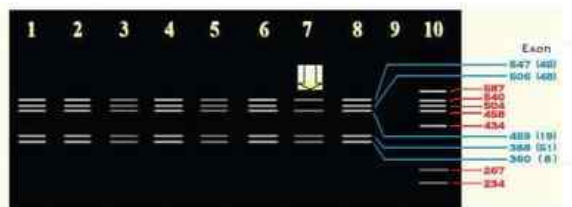


Fig. Results of multiplex PCR in a patient with Duchenne

Muscular Dystrophy Inverse PCR
Hernandez-Rodriguez et al., 2000;
Hernandez-Rodriguez &
Restrepo, 2002 (www.intechopen.com)

Lecture 169

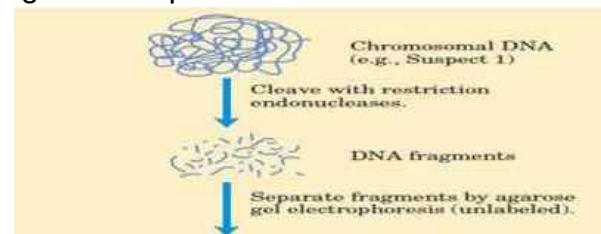
RAPD:

□ RAPD-Random amplification of polymorphic DNA is a type of PCR

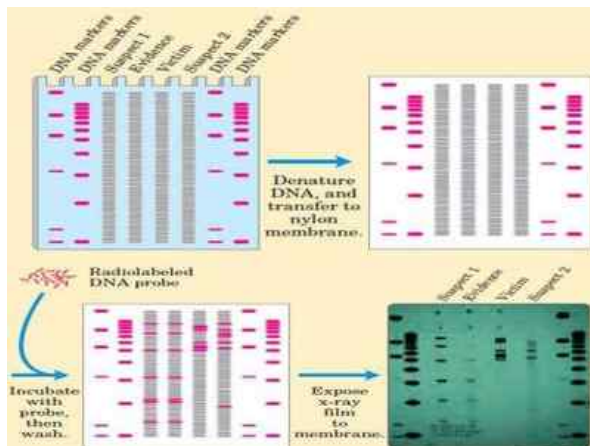
Lecture 170

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



Lecture 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.

Lecture 172

Applications of PCR

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

PCR-Gene cloning and expression

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

Lecture 173

- **PCR-Medicine:** PCR has major impact on medicine especially in the field of clinical microbiology/diagnosis

Molecular tools have also allowed to perform prenatal genetic diagnosis.

Lecture 174

PCR-Forensic sciences: Forensic science is the application of scientific procedures to solve criminal & legal matters

- Molecular methods are used to establish 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

Lecture 175

PCR-Agricultural sciences and environment

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

Lecture 176

PCR-Molecular paleontology

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

Lecture 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

Lecture 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

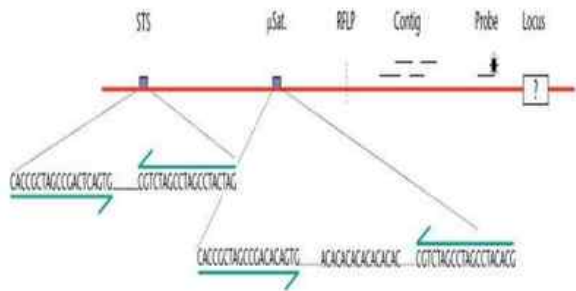


Figure. Markers used in genome mapping

(Paul H Dear, 2001,

Encyclopedia of life sciences)

Lecture 179

Methods for genome mapping

Genetic linkage mapping:

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

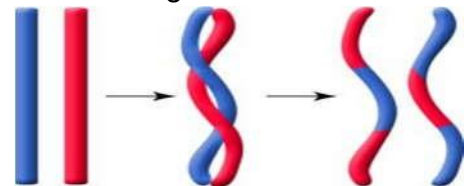


Figure. Crossing over in chromosomes

Source: Genomic News Network (GNN)

Lecture 180

Physical mapping

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

Lecture 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.

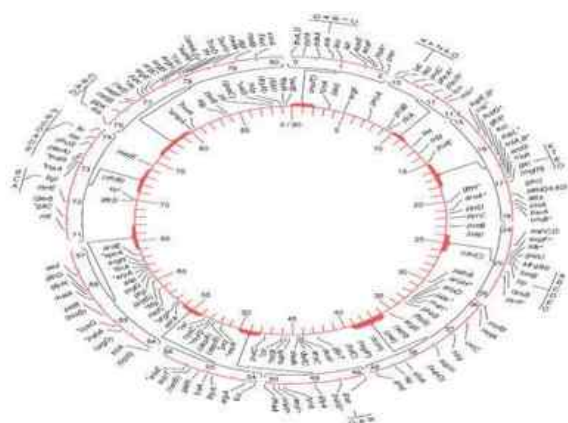


Figure. The 1963 map of *E. coli*

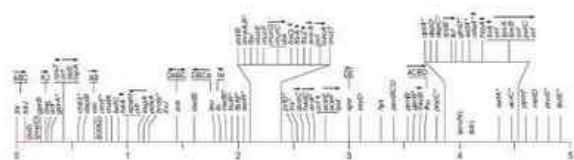


Figure. Linear scale drawing of a 5-minute section of the 100-minute 1990 *E. coli* linkage map

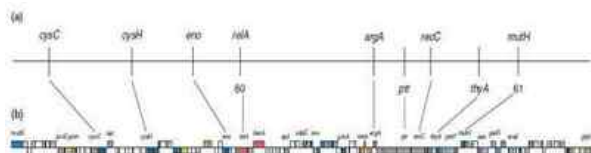
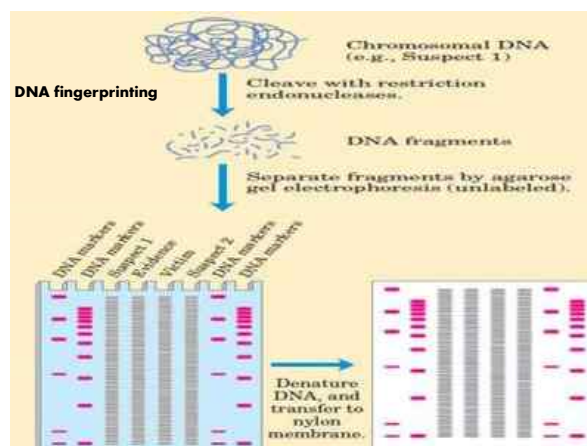
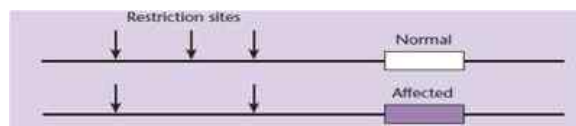


Figure. Correlation of the genetic and physical maps

Lecture 182

The use of RFLPs in physical maps:

RFLP-Restriction fragment length polymorphism exploits variations in homologous DNA sequences. Botstein et al. (1980) were the first to recognize that DNA probes that target RFLPs can be used for mapping.



Lecture 183

STS in physical maps

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

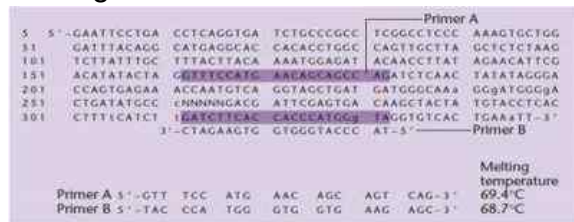


Figure. Examples of a sequence-tagged site (STS)

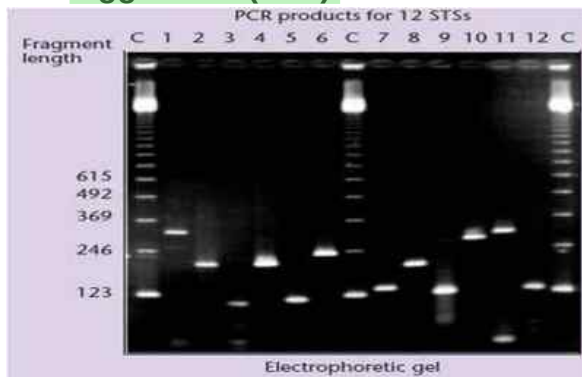


Figure. Confirmation that an STS is a unique sequence on the genome

Lecture 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.

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

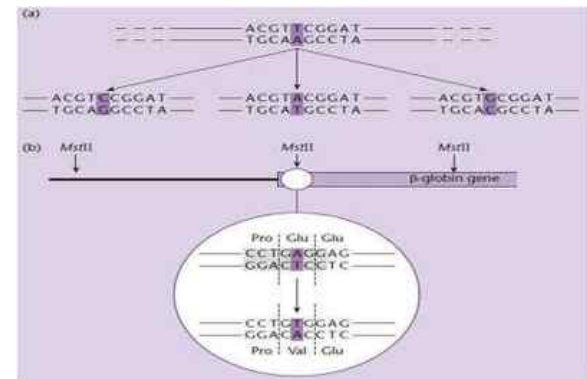


Figure. Examples of single nucleotide polymorphisms

Lecture 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.

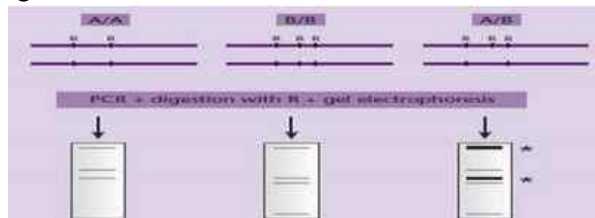


Figure. Generation and visualization of CAPS markers

Lecture 186

AFLPs detection in the absence of sequence information

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

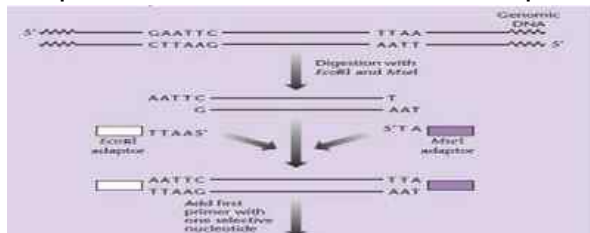
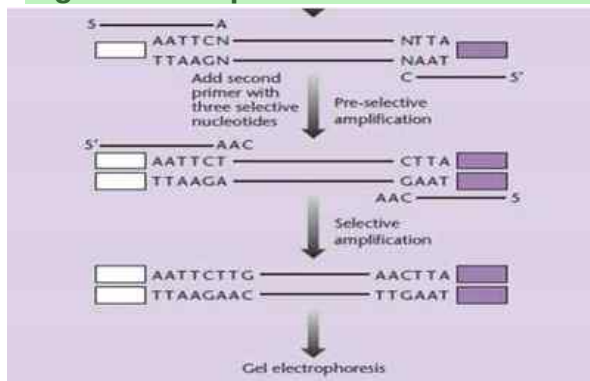


Figure. Principle of the AFLP method



Lecture 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.

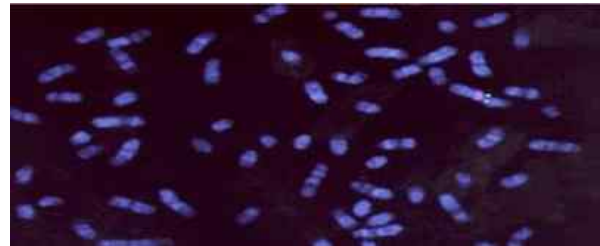


Figure. Dual-color fluorescence in situ hybridization

Lecture 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.

Lecture 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.

Lecture 190

Map integration

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

Lecture 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.

Lecture 192

High-throughput sequencing

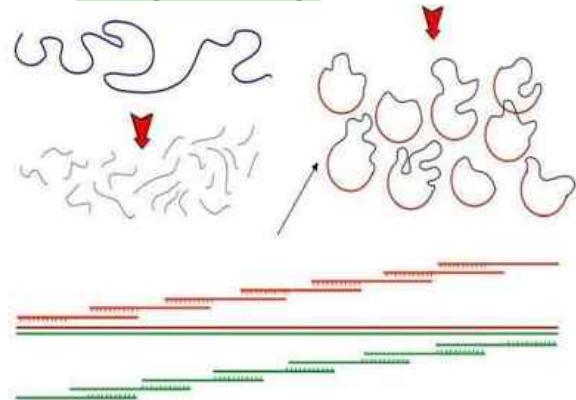
- Recent advances in DNA sequencing have made it possible to sequence data very rapidly and at a substantially lower cost
- It mainly describes a number of different modern sequencing technologies to sequence genome

Lecture 193

Shotgun sequencing

- Shotgun sequencing is used for sequencing long DNA strands
- In this strategy, DNA is shredded into smaller fragments that can be sequenced individually and then reassembled into original order.

Shotgun library



Lecture 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.

Lecture 195

Orthologs and paralogs

- Orthologs are homologous genes in different organisms that encode proteins with the same function and which have evolved by direct vertical descent.

Lecture 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*)

Lecture 197

Comparative genomics of organelles

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

Lecture 198

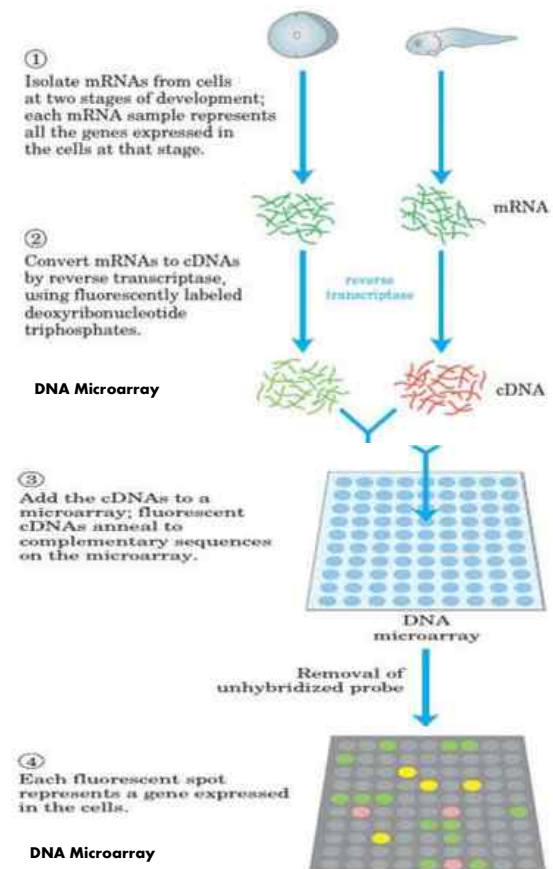
Comparative genomics of eukaryotes

- 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

Lecture 199

DNA Microarrays

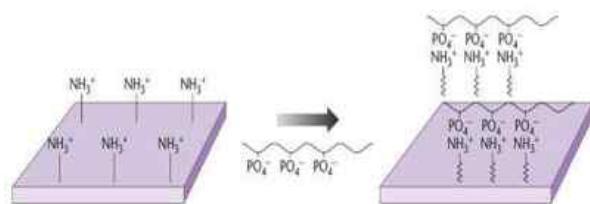
- Microarrays are used to study the expression of many genes at once. It involves the immobilization of thousands of gene sequences on glass slide that are then detected with complementary base pairing between the sample and the gene sequences on the chip



Enlarged image of a DNA Microarray

DNA Microarrays

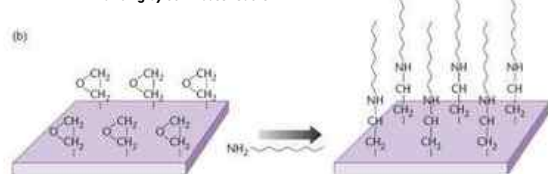
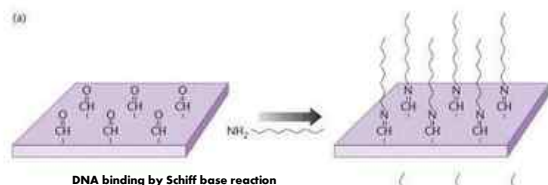
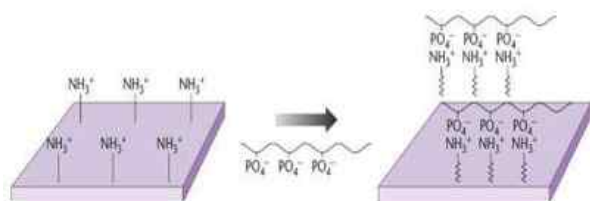
- Two major types of DNA array are used in expression analysis
 - Spotted DNA arrays
 - Printed oligonucleotide chips



Lecture 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

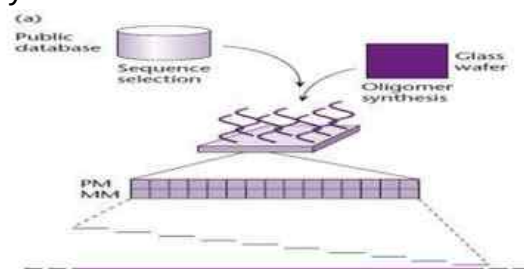


b)- Binding of DNA molecules to epoxy-derivatized surfaces

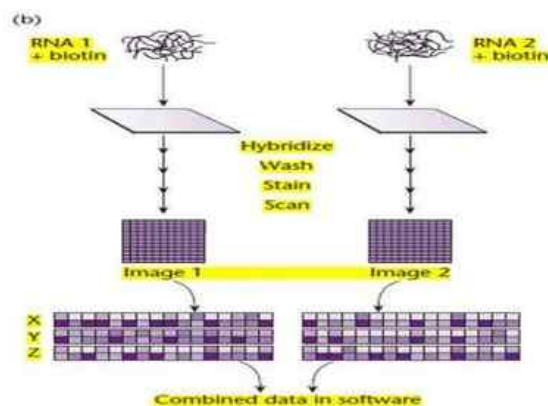
Lecture 201

Oligonucleotide chips

□ Oligonucleotide chips are manufactured by in situ oligonucleotide synthesis



Principle of chip manufacture



Principle of chip hybridization

Lecture 202

Applications of microarrays

Microbial gene expression analysis

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

Lecture 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

Lecture 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

Knock ins

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

Lecture 204

Phage display

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

Lecture 205

Functional genomics and proteomics

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

Lecture 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

Lecture 208

siRNA technology

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

Lecture 209

Applications of siRNA

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