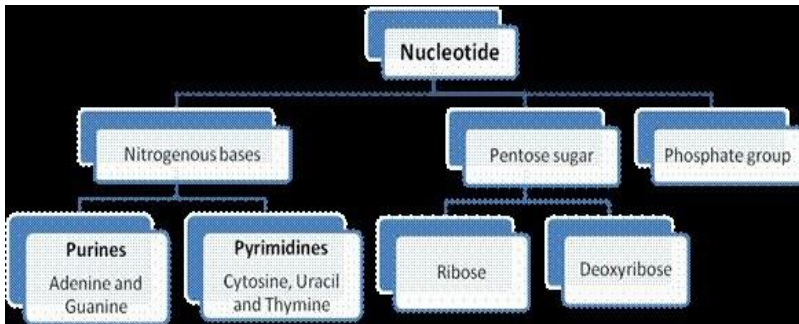


## MOLECULAR BASIS OF INHERITANCE

DNA (Deoxyribonucleic Acid) and RNA (Ribonucleic Acid) are two types of nucleic acid found in living organisms. DNA acts as genetic material in most of the organisms. RNA also acts as genetic material in some organisms as in some viruses and acts as messenger. It functions as adapter, structural, and in some cases as a catalytic molecule

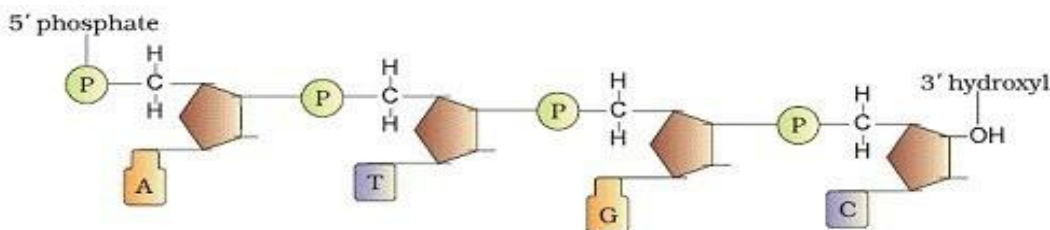
The DNA - it is a long polymer of deoxyribonucleotides. A pair of nucleotide is also known as base pairs. Length of DNA is usually defined as number of nucleotides present in it. Escherichia coli have  $4.6 \times 10^6$  bp and haploid content of human DNA is  $3.3 \times 10^9$  bp.

### Structure of Polynucleotide Chain



- A nucleotide has three components – a nitrogenous base, a pentose sugar (ribose in case of RNA, and deoxyribose for DNA), and a phosphate group. There are two types of nitrogenous bases – Purines (Adenine and Guanine), and Pyrimidines (Cytosine, Uracil and Thymine).

Cytosine is common for both DNA and RNA and Thymine is present in DNA. Uracil is present in RNA at the place of Thymine.



## A polynucleotide chain

A nitrogenous base is linked to pentose sugar with N-glycosidic linkage to form a nucleoside. When phosphate group is linked 5'-OH of a nucleoside through phosphoester linkage nucleotide is formed. Two nucleotides are linked through 3'-5' phosphodiester linkage to form dinucleotide. More nucleotide joins together to form polynucleotide.

In RNA, nucleotide residue has additional -OH group present at 2'-position in ribose and uracil is found at the place of Thymine.

### Structure differences

DNA	RNA
(a) The sugar present in DNA is 2-deoxy-D - (-) -ribose.	(a) The sugar present in RNA is D- (-)- ribose.
(b) DNA contains cytosine and thymine as pyrimidine bases and guanine and adenine as purine bases.	(b) RNA contains cytosine and uracil as pyrimidine bases and guanine and adenine as purine bases.
(c) DNA has double strand $\alpha$ -helix structure.	(c) RNA has a single stranded $\alpha$ -helix structure.
(d) DNA molecules are very large their molecular mass may vary from $6 \times 10^6 - 16 \times 10^6 u$	(d) RNA molecules are comparatively much smaller with molecular mass ranging from 20,000 - 40,000.

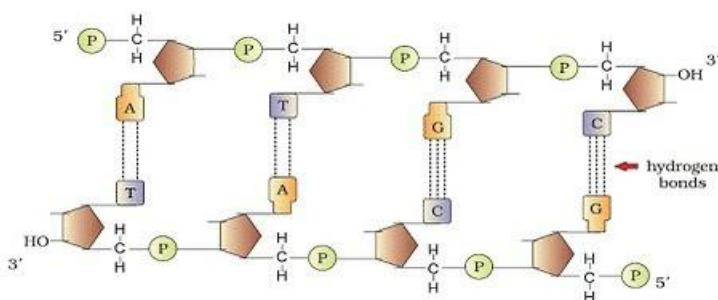
## Functional differences

(a) DNA has unique property of replication.	(a) RNA usually does not replicate.
(b) RNA controls the transmission of hereditary effects.	(b) RNA controls the synthesis of proteins.

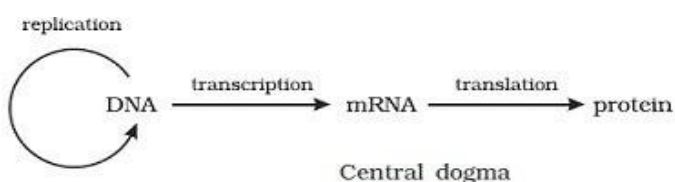
Double Helix Model for Structure of DNA-James Watson and Francis Crick, based on X-ray diffraction data produced by Wilkin and Rosalind proposed this model of DNA.

The silent features of this model are-

- DNA is made of two polynucleotide chains in which backbone is made up of sugar-phosphate and bases projected inside it.
- Two chains have anti-parallel polarity. One 5'→3' and with 3'→5'.
- The bases in two strands are paired through H-bonds. Adenine and Thymine forms double hydrogen bond and Guanine and Cytosine forms triple hydrogen bonds.
- Two chains are coiled in right handed fashion. The pitch of helix is 3.4 nm and roughly 10 bp in each turn.
- The plane of one base pair stacks over the other in double helix to confer stability.



- Francis Crick proposed the Central dogma in molecular biology, which states that the genetic information flows from DNA → RNA → Protein.

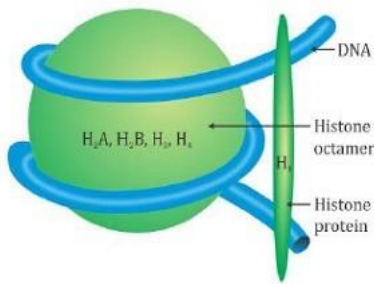


### Packing of DNA helix-

In prokaryotes, well defined nucleus is absent and negatively charged DNA is combined with some positively charged proteins called nucleoids.

In eukaryotes, histones, positively charged protein organized to form 8 molecules unit called histone octamer. Negatively charged DNA is wrapped around the histone octamer to form nucleosome. . Histones are rich in the basic amino acid residues lysines and arginines. Both the amino acid residues carry positive charges in their side chains.

- Single nucleosome contains about 200 base pairs. Chromatin is the repeating unit of nucleosome.



Structure of Nucleosome

- In nucleus, some region of chromatin are loosely packed (and stains light) and are referred to as euchromatin. The chromatin that is more densely packed and stains dark are called as Heterochromatin. Euchromatin is transcriptionally active chromatin, whereas heterochromatin is inactive.

## The search for Genetic Material

Transforming principle – Frederick Griffith in 1928 conducted experiment on bacteria *Streptococcus pneumoniae* (bacterium responsible for pneumonia). There are two types of strain of this bacteria, some produce smooth shiny colonies (S) and others produce rough colonies (R). Mice infected with the S strain (virulent) die from pneumonia infection but mice infected with the R strain do not develop pneumonia.

S strain → Inject into mice → Mice die

R strain → Inject into mice → Mice live

S strain (heat-killed) → Inject into mice → Mice live

S strain (heat-killed) + R strain (live) → Inject into mice → Mice die

Griffith concluded that R strain bacteria have somehow transformed by heat killed S strain bacteria. Some transforming principles transferred from S strain to R strain and enabled the R strain to synthesise a smooth polysaccharide coat and become virulent. This must be due to the transfer of the genetic material.

## Biochemical Characterisation of Transforming Principle

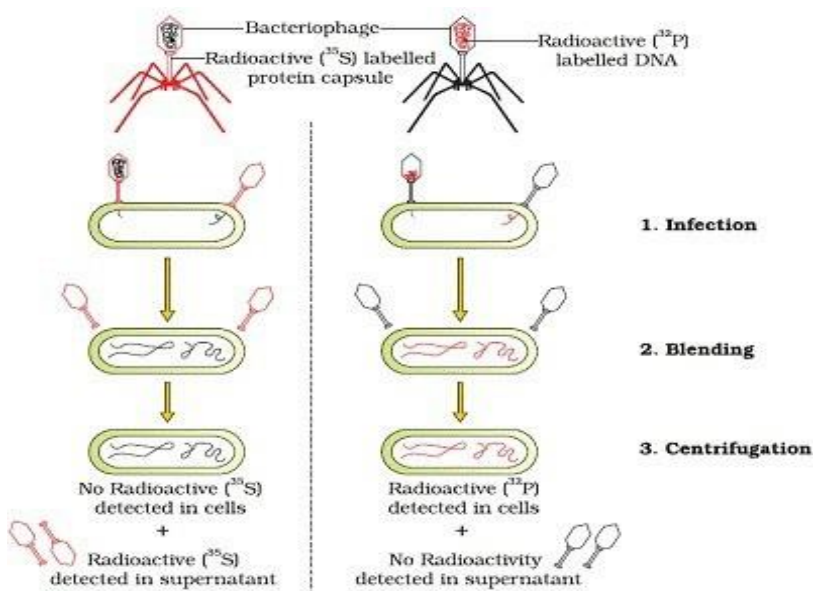
- Oswald Avery, Colin MacLeod and Maclyn McCarty worked out to determine the biochemical nature of transforming principle of Griffith.
- They purified biochemicals (proteins, DNA, RNA, etc.) from the heat-killed S cells to see which ones could transform live R cells into S cells. They discovered that DNA alone from S bacteria caused R bacteria to become transformed. So, they concluded that DNA is the genetic material.

## Experimental proof that DNA is the genetic material

Alfred Hershey and Martha Chases (1952) worked with virus that infect bacteria called bacteriophages.

- In one preparation, the protein part was made radioactive and in the other, nucleic acid (DNA) was made radioactive. These two phage preparations were allowed to infect the culture of E.coli. Soon after infection, before lysis of cells, the E.coli cells were gently agitated in a blender, to loosen the adhering phage particles and the culture was centrifuged.
- The heavier infected bacterial cells pelleted to the bottom and the lighter viral particles were present in the supernatant. It was found that when bacteriophage containing radioactive DNA was used to infect E.coli, the pellet contained radioactivity.
- If bacteriophage containing radioactive protein coat was used to infect E.coli, the supernatant contained most of the radioactivity.

His experiment shows that protine does not enter the bacterial cell and only DNA is the genetic material.



### Properties of Genetic Material:

- a) It should be able to generate its replica (replication)
- b) It should chemically and structurally be stable.
- c) It should provide the scope for slow changes (mutation) that are required for evolution.
- d) It should be able to express itself in the form of 'Mendelian Characters'.

- DNA is chemically less reactive but structurally more stable as compared to RNA. So, DNA is better genetic material.
- RNA used as genetic material as well as catalyst and more reactive so less stable. Therefore, DNA has evolved from RNA.

### Replication of DNA

Watson and Crick suggested that two strands of DNA separate from each other and act as template for synthesis of new complementary strands. After the completion of replication each DNA molecule would have one parental and one newly synthesised strand, this method is called semiconservative replication.

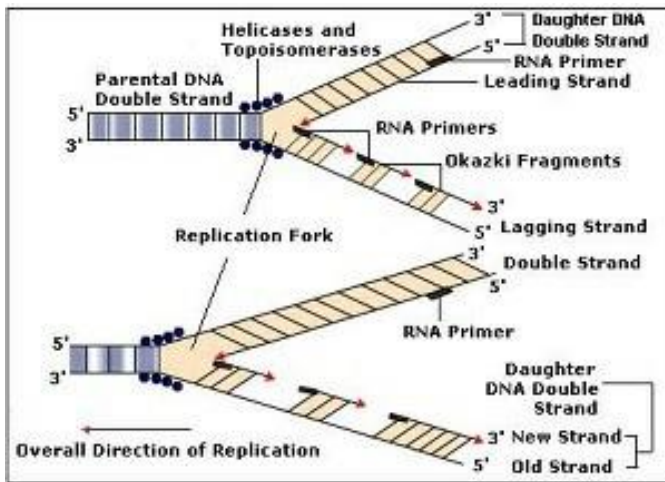
- Messelson and Stahl's shows experimental evidence of semiconservative replication by growing *E. coli* on nutrient media containing nitrogen salts ( $^{15}\text{NH}_4\text{Cl}$ ) labeled with radioactive  $^{15}\text{N}$ .
  - $^{15}\text{N}$  was incorporated into both the strands of DNA and such a DNA was heavier than the DNA obtained from *E. coli* grown on a medium containing  $^{14}\text{N}$ . Then they transferred the *E. coli* cells on to a medium containing  $^{14}\text{N}$ .
  - After one generation, when one bacterial cell has multiplied into two, they isolated the DNA and evaluated its density. Its density was intermediate between that of the heavier  $^{15}\text{N}$ -DNA and the lighter  $^{14}\text{N}$ -DNA.
  - This is because during replication, new DNA molecule with one  $^{15}\text{N}$ -old strand and a complementary  $^{14}\text{N}$ -new strand was formed (semi-conservative replication) and so its density is intermediate between the two.

Replication : Replication of DNA require Enzyme DNA polymerase that catalyse the polymerisation in one strand  $5' \rightarrow 3'$  only after unwinding with the help of Helicase enzyme . So, replication in one strand is continuous and other strand it is discontinuous to synthesise okazaki fragments that are joined together by enzyme DNA ligase.  $\rightarrow$

### Differences between leading and Lagging strands for DNA replication.

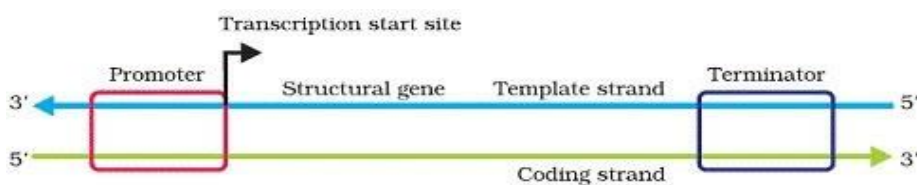
Characters	Leading strand	Lagging strand
1. Fragments	It is formed continuously as single fragment.	In the beginning it is formed in the form of small fragments called okazaki segments.
2. RNA primer	It requires only one primer to initiate the growth.	Every fragment requires separate RNA primer to initiate.
3. DNA ligase	Not required.	Required to join DNA fragments.
4. Direction of growth	$5' \rightarrow 3'$	Of complete strand it is $3' \rightarrow 5'$ . However for okazaki fragments it is $5' \rightarrow 3'$ .





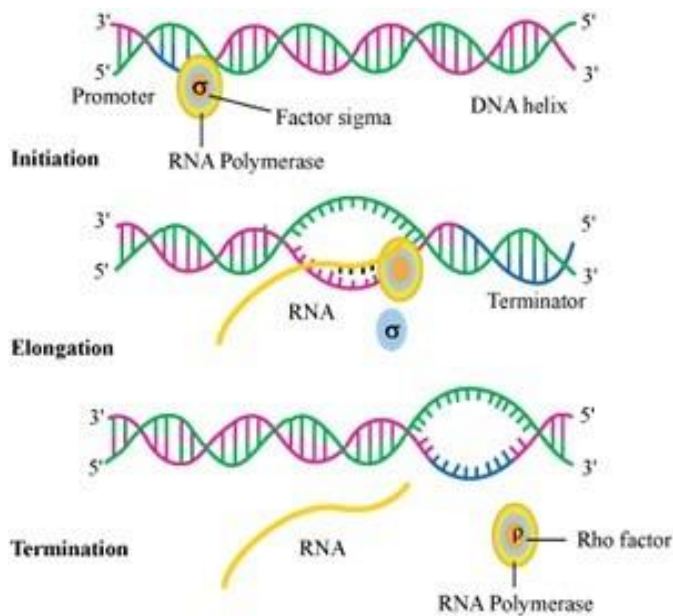
## Transcription

- It is the process of copying genetic information from one strand of DNA into RNA. In transcription only one segment of DNA and only one strand is copied in RNA. The Adenosine forms base pair with Uracil instead of Thymine.
- Transcription of DNA includes a promoter, the structural gene and a terminator. The strands that has polarity 3' to 5' act as template and called template strand and other strand is called coding strand.



Template Stranc	Coding Strand
It is a DNA strand with 3' → 5' Polarity.	DNA Strand with 5' → 3' Polarity
Acts as template for transcription and codes for RNA	Does not code for any region of RNA during transcription.

Promoter is located at 5' end and that bind the enzyme RNA polymerase to start transcription. Sigma factor also help in initiation of transcription .The terminator is located at 3' end of coding strand and usually defines the end of transcription where rho factor will bind to terminate transcription.



Exons are those sequences that appear in mature and processed RNA. Exons are interrupted by introns. Introns do not appear in mature and processed RNA.

- In eukaryotes, there are three different RNA polymerase enzymes I, II and III, they catalyse the synthesis of all types of RNA.

RNA polymerase I – rRNAs

RNA polymerase II - mRNA

RNA polymerase III – tRNA

The m-RNA provide the template, t-RNA brings the amino acids and read the genetic code, the r-RNA play structural and catalytic role during translation.

DNA Replication	RNA Transcription
Two new molecules of double-stranded DNA are produced.	One new molecule of single-stranded RNA is produced.
Adenine on one strand binds to thymine on the new DNA strand being created.	Adenine on DNA binds to uracil on the new RNA strand being created.

The entire chromosome is replicated.	Only a small portion of the DNA molecule is transcribed to RNA, and this varies based on the cell's needs at the time.
Enzymes: DNA polymerase	Enzymes: RNA polymerase
Occurs in nucleus.	Occurs in nucleus.

The primary transcript contains both exon and intron and is non-functional. It undergoes the process of splicing in which introns are removed and exons are joined in a defined order.

The hnRNA (heterogeneous nuclear RNA) undergo additional processing called as capping and tailing. In capping in unusual nucleotide (methylguanosine triphosphate) to the 5' end of hnRNA. In tailing polyadenylate tail is added at 3' end in a template at independent manner.

Genetic Code : Genetic Code is the relationship of amino acids sequence in a polypeptide and nucleotide/base sequence in mRNA. It directs the sequence of amino acids during synthesis of proteins.

George Gamow suggested that genetic code should be combination of 3 nucleotides to code 20 amino acids.

H.G. Khorana developed chemical method to synthesising RNA molecules with defined combination of bases.

Marshall Nirenberg's cell free system for protein synthesis finally helped the code to be deciphered.

Salient features of Genetic Code are-

- i. The code is triplet. 61 codons code for amino acids and 3 codons do not code for any amino acids called stop codon (UAG, UGA and UAA).
- ii. Codon is unambiguous and specific, code for one amino acid.

- iii. The code is degenerate. Some amino acids are coded by more than one codon.
- iv. The codon is read in mRNA in a contiguous fashion without any punctuation.
- v. The codon is nearly universal. AUG has dual functions. It codes for methionine and also act as initiator codon.

### Mutations and Genetic code

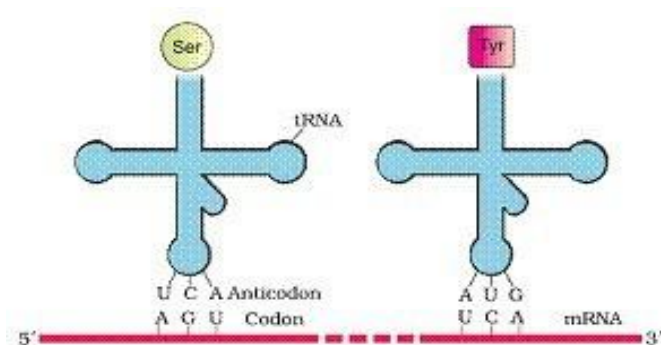
A change of single base pair (point mutation) in the 6th position of Beta globin chain of Haemoglobin results due to the change of amino acid residue glutamate to valine. These results into diseased condition called sickle cell anaemia.

Insertion and deletion of three or its multiple bases insert or delete one or multiple codons hence one or more amino acids and reading frame remain unaltered from that point onwards. Such mutations are called frame-shift insertion or deletion mutations.

### tRNA– the Adapter Molecule

The t-RNA called as adaptor molecules. It has an anticodon loop that has bases complementary to code present on mRNA and also has an amino acid acceptor to which amino acid binds. t-RNA is specific for each amino acids.

The secondary structure of t-RNA is depicted as clover-leaf. In actual structure, the t-RNA is a compact molecule which look like inverted L.

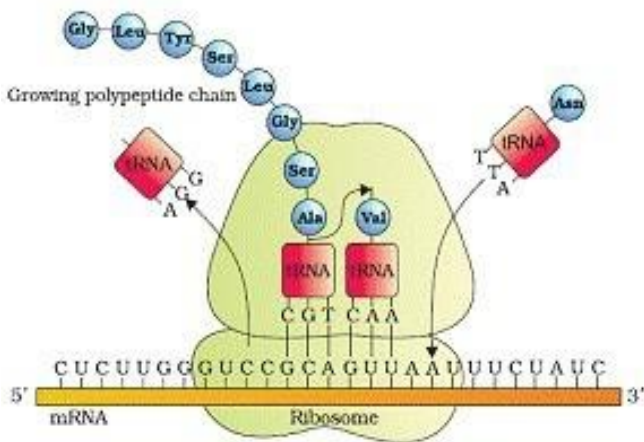


**Translation process:** Translation is the process of polymerisation of amino acids to form a polypeptide. The order and sequence of amino acids are defined by the sequence of bases in the mRNA. Amino acids are joined by peptide bonds. It involved following steps-

a) Charging of t-RNA.

b) Formation of peptide bonds between two charged tRNA.

- The start codon is AUG. An mRNA has some additional sequence that are not translated called untranslated region (UTR).
- For initiation ribosome binds to mRNA at the start codon. Ribosomes moves from codon to codon along mRNA for elongation of protein chain. At the end release factors binds to the stop codon, terminating the translation and release of polypeptide form ribosome.

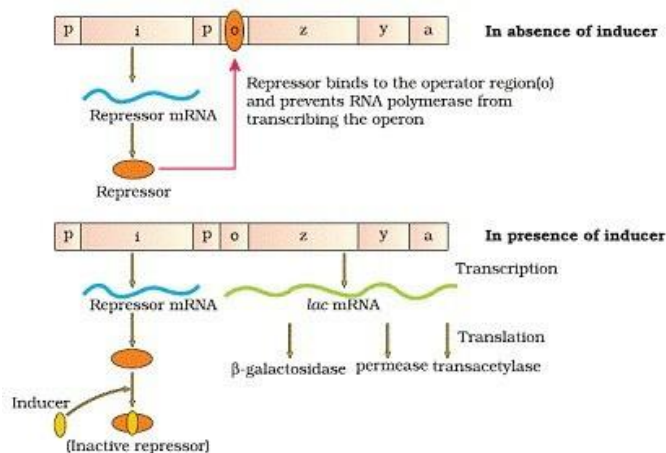


### Regulation of Gene Expression:

All the genes are not needed constantly. The genes needed only sometimes are called regulatory genes and are made to function only when required and remain non-functional at other times. Such regulated genes, therefore required to be switched 'on' or 'off' when a particular function is to begin or stop.

### The Lac Operon

Lac operon consists of one regulatory gene (i) and three structural genes (y, z and a). Gene i code for the repressor of the lac operon. The z gene code for beta-galactosidase, that is responsible for hydrolysis of disaccharide, lactose into monomeric units, galactose and glucose. Gene y code for permease, which increases permeability of the cell. Gene a encode for transacetylase.



Lactose is the substrate for enzyme beta-galactosidase and it regulates switching on and off of the operon, so it is called inducer.

Regulation of Lac operon by repressor is referred as negative regulation. Operation of Lac operon is also under the control of positive regulation.

Human Genome Project was launched in 1990 to find out the complete DNA sequence of human genome using genetic engineering technique and bioinformatics to isolate and clone the DNA segment for determining DNA sequence.

### Goal of HGP-

- Identify all the genes (20,000 to 25,000) in human DNA.
- Determine the sequence of the 3 billion chemical base pairs that make up human DNA.
- Store this information in data base.
- Improve tools for data analysis.
- Transfer related information to other sectors.
- To address the legal, ethical and social issues that may arise due to project.

- The project was coordinated by the US Department of Energy and the National Institute of health.
- The method involved the two major approaches- first identifying all the genes that express as RNA called Express sequence tags(EST).The second is the sequencing the all set of genome that contained the all the coding and non-coding sequence called sequence Annotation.

### Salient features of Human Genome:

- a) The human genome contains 3164.7 million nucleotide bases.
- b) The average gene consists of 3000 bases, but sizes vary greatly, with the largest known human gene being dystrophin at 2.4 million bases.
- c) Less than 2 per cent of the genome codes for proteins.
- d) Repeated sequences make up very large portion of the human genome.
- e) Repetitive sequences are stretches of DNA sequences that are repeated many times, sometimes hundred to thousand times.
- f) Chromosome 1 has most genes (2968), and the Y has the fewest (231).
- g) Scientists have identified about 1.4 million locations where single base DNA differences (SNPs – single nucleotide polymorphism) occur in humans.

DNA finger printing is a very quick way to compare the DNA sequence of any two individual. It includes identifying differences in some specific region in DNA sequence called as repetitive DNA because in this region, a small stretch of DNA is repeated many times.

Depending upon the base composition, length of segment and number of repetitive units satellite DNA is classified into many categories.

Polymorphism in DNA sequence is the basis for genetic mapping of human genome as well as fingerprinting.

The technique of fingerprinting was initially developed by Alec Jeffrey. He used a satellite DNA as probe to so high polymorphism was called Variable Number of Tendon Repeats (VNTR).

