MOLECULAR BASIS OF INHERITANCE

- Genetic material is that substance which controls the formation and expression of traits in an organism and can replicate and pass on from a cell to its daughter cell or from one generation to next.
- The inheritance patterns and the genetic basis of such patterns was not clear at the time of Mendel. Over the next hundred years, the nature of the genetic material was investigated culminating in the realisation that deoxyribonucleic acid (DNA) is the genetic material or the molecular basis of inheritance for the majority of organisms.

DNA AS GENETIC MATERIAL

- The most conclusive evidences in support of DNA as the genetic material came from the analysis of **transformation of bacteria, mode of infection of bacteriophages and conjugation of bacteria.**
- **Frederick Griffith** conducted a series of experiments with bacteria *Streptococcus pneumoniae* (or *Diplococcus pneumoniae.*)
- Griffith conducted a series of experiments as given:
  - Injected virulent S-III strains into mice, the mice developed pneumonia and died.
  - Injected non-pathogenic R-II strains into mice, the mice did not get the disease and survived.
  - Injected heat killed S-III strains into mice, the mice did not get the disease and remained alive.
  - Injected heat killed S-III strains along with live R-II strains, the mice suffered from pneumonia and died.
- In the last experiment, Griffith did not inject living S-III strains into mice but living S-III strains appeared in the blood and caused the death of mice. Griffith concluded that there was some factor in heat killed S-III strains that **transformed** live R-II strains into live S-III strains. This must be due to the transfer of the genetic material. However, the biochemical nature of genetic material was not defined from his experiments.
- In 1944, **Avery, McCarty and MacLeod** fractionated the killed S-type bacteria into three components-**DNA, carbohydrate and protein**. They also discovered that protein-digesting enzymes (proteases) and RNA-digesting enzymes (**RNases**) did not affect transformation, so the transforming substance was not a protein or RNA. Digestion with **DNase** did not inhibit transformation, suggesting that the **DNA caused the transformation.** They concluded that DNA is the hereditary material.
- The unequivocal proof that DNA is the genetic material came from the experiments of **Alfred Hershey and Martha Chase (1952).** They worked with viruses that infect bacteria called bacteriophages to discover whether it was protein or DNA from the viruses that entered the bacteria.
- They incorporated radioactive isotope of **phosphorus (^{32}P) into phage DNA** in a phage culture and that of **sulphur (^{35}S) into proteins** of a separate phage culture. These phage types were used independently to infect the bacterium *Escherichia coli.* After sometime, this mixture was agitated on a blender to separate the empty phage capsids from the surface of bacterial cell. When ^{35}S was used, all radioactive material was limited to empty viral protein coats.
- Bacteria which was infected with viruses that had radioactive DNA were radioactive indicating that DNA was the genetic material that passed from the virus to the bacteria.
- These results indicated that the DNA of the bacteriophage and not the protein enters the host, where viral replication takes place. Therefore, DNA is the genetic material of bacteriophage.

DNA

- DNA (deoxyribonucleic acid) is a long **polymer of deoxyribonucleotides**. Chemically, DNA is composed of three components **a pentose sugar, phosphoric acid, and four types of nitrogenous bases**. Pentose sugar in DNA is deoxyribose sugar.
- The four nitrogenous bases belong to two separate groups, (i) purines which are two-ringed nitrogen compounds and include adenine and guanine, and (ii) pyrimidines which are formed of one ring only and include cytosine, and thymine.
- **James Watson** and **Francis Crick** proposed that DNA consists of two strands, which are **helically coiled**. The two strands are said to be complementary. The strands of DNA are said to be antiparallel, *i.e.*, one in 5′--> 3′ direction and the other in 3′ --> 5′ direction.
- There are two hydrogen bonds between A and T and three between G and C. The stacking of bases creates two types of grooves called **major** and **minor grooves**. Each turn accommodates **10 bases**. The two chains are coiled in a **right-handed fashion**. The pitch of the helix is **3.4 nm**.

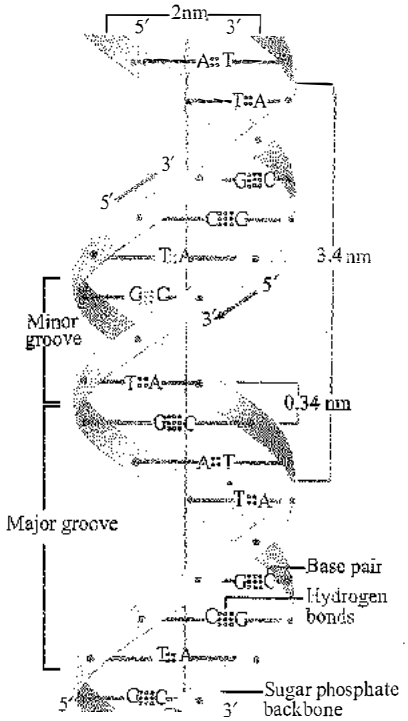


Fig.: Double helical DNA strand.

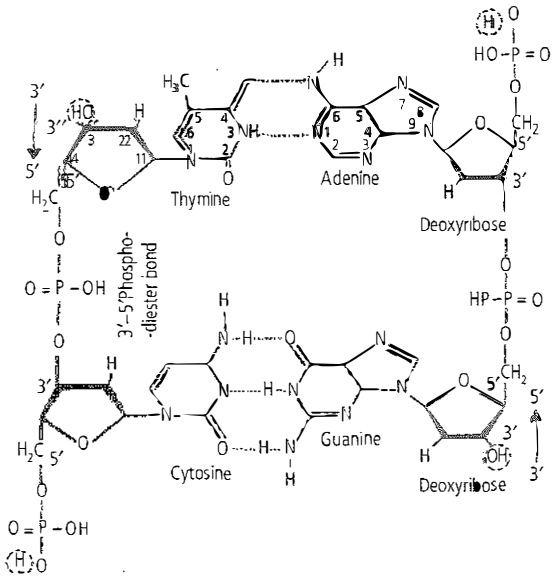


Fig.: Arrangement of various constituents of DNA duplex.

- DNA duplex model proposed by Watson and Crick is right handed spiral and is called **B-DNA**. Another right handed duplex model is **A-DNA**. Here, a single turn of helix has 11 base pairs. **C-DNA** has **9 base pairs** per turn of spiral while in **D-DNA** the number is only 8 base pairs. Both are right handed. **Z-DNA is left-handed double helix with zig-zag backbone**, altemate purine and pyrimidine bases, single turn of 45 Å length with **12 base pairs** and a single groove.
- In 1950, **Erwin Chargaff** formulated important generalizations about DNA structure, these generalization are called **Chargaff's rules** in his honour. They are summarized as follows :
  - The DNA molecule, irrespective of its source, always has the A – T base pairs equal in number to the G – C base pairs.
  - The purines and pyrimidines are always in equal amounts, *i.e.*, A + G = T + C.
  - The amount of adenine is always equal to that of thymine, and the amount of guanine is always equal to that of cytosine, *i.e.*, A = T and G = C. However, amount of A + T is not necessarily equal to G + C.
  - The base ratio A + T / G + C may vary from one species to another, but is constant for a given species.
- In prokaryotes, where nucleus is absent, the DNA is not scattered throughout the cell. DNA (being negatively charged) is held with some non-histone proteins (that have positive charges) in a region termed as '**nucleoid**'. In eukaryotes, this organisation is much more complex. There is a set of positively charged, basic proteins called **histones**. Histones are rich in the basic amino acid residues **lysines** and **arginines**. To assist in packaging of very long DNA molecules (approx. 2.2 mts in humans) in a very small cell, histones are organized to form a unit of eight molecules called **histone octamer**. Histones are of five types- H_1, H_2A, H_2B, H_3, H_4. The negatively charged DNA is wrapped around the positively charged histone octamer to form a structure called **nucleosome**.
- Nucleosomes constitute the repeating unit of a structure in nucleus called **chromatin**. Some regions of chromatin are loosely packed (and stain light) and are referred to as **euchromatin**. The chromatin that is more densely packed and stains dark is called as **heterochromatin**.
- **Euchromatin is transcriptionally active chromatin, whereas heterochromatin is inactive.**

RNA

- RNA (ribonucleic acid) is a non-hereditary nucleic acid except in some viruses (**retroviruses**). It is a **polymer of ribonucleotides** and is made up of ribose sugar, phosphoric acid and nitrogenous bases (A, U, C, G). The chemical composition of RNA differs from DNA in two respects (i) the pentose sugar in RNA is ribose sugar while it is deoxyribose sugar in DNA, (ii) Thymine in DNA is replaced by uracil in RNA. All other nitrogenous bases are same in both.

Types of RNA

- The RNA may be mainly of two types – genetic RNA and non-genetic RNA. **Genetic RNA** is seen in most of the plant viruses and some animal viruses, *e.g.*, in **TMV, polio virus, influenza virus, etc., RNA acts as genetic material.**

Table : Different RNA viruses and the nature of genetic RNA associated with them

	Virus	Types of RNA
1.	**Plant viruses**	
	TMV.	Single stranded.
	Wound tumor	Double stranded
2.	**Animal viruses**	
	Influenza, Rous sarcoma, poliomyelitis.	Single stranded
	Reovirus	Double stranded
3.	**Bacteriophages**	
	MS_2, F_2, R_{17}	Single stranded

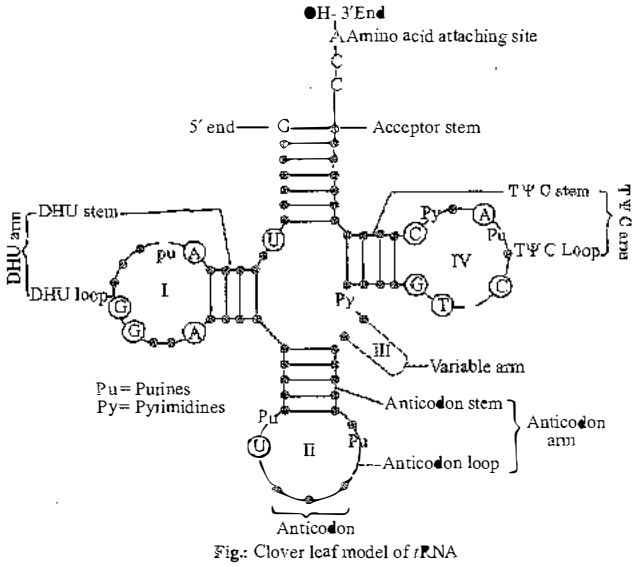


Fig.: Clover leaf model of *t*RNA

- On the basis of molecular size and function, three main forms of non genetic RNA are – *m*RNA, *t*RNA and *r*RNA.
- *m*RNA constitutes about 3.5% of cellular RNA; *t*RNA is about 15% and *r*RNA about 80%.

- **Messenger ribonucleic acid** (*m*RNA) or informational RNA or template RNA is a molecule of RNA that is transcribed from a gene and then translated by ribosomes in order to manufacture protein. The name *m*RNA was given by **Jacob and Monod** (1961), because it is responsible for carrying the genetic information present in DNA, *i.e.*, acts as a **messenger of informations contained in DNA**.
- When a particular gene codes for a *m*RNA strand, it is said to be **monocistronic or monogenic**, *e.g.*, eukaryotes. When several genes (cistrons) are transcribed into a single *m*RNA molecule, it is described as **polycistronic or polygenic**, *e g* , prokaryotes. Monocistronic *m*RNA codes for one complete protein molecule while polycistronic *m*RNA codes for a number of protein molecules.
- *t*RNA or transfer RNA is also known as **soluble (s) RNA, acceptor RNA or adaptor RNA**. *t*RNAs are smallest, bearing 70-80 nucleotides. **Holley (1965)** reported the sequence of an alanine *t*RNA from yeast and proposed two-dimensional structure of *t*RNA (**clover leaf model**). Three dimensional (**L-shaped**) structure of *t*RNA was proposed by **Klug (1974)**.
- In *t*RNA five regions that are single stranded are **AA (amino acid) binding site, T Ψ C loop, DHU loop, extra arm and anticodon loop**.
- **AA - binding site** lies opposite to the anticodon site. *t*RNA molecules have unpaired (single stranded) CCA–OH sequence at the 3′ end. This is called amino acid attachment site.
- **T Ψ C loop** contains pseudouridine. The loop is the site for attaching to ribosomes. **DHU loop** contains **dihydrouridine**. It is binding site for aminoacyl synthetase enzyme.
- **Extra arm** is a variable site or loop which lies between T Ψ C loop and anticodon. The exact role of extra arm is not known. **Anticodon loop** is made up of three nitrogen bases for recognising and attaching to the codon of *m*RNA.
- *t*RNA helps to transport amino acids from the surrounding cytoplasm to the site of protein synthesis.
- **Ribosomal RNA** is a component of the ribosomes, the protein synthetic factories in the cell. It is formed in nucleolus. ***r*RNA is the most stable type of RNA**.
- Depending upon sedimentation coefficient **eukaryotic ribosomes** contain four different *r*RNA molecules : **18S, 5.8S, 28S and 5S *r*RNA**. Prokaryotic ribosomes have three types of *r*RNAs: **23S, 16S and 5S.** Among these **28S, 5.8S and 5S** in eukaryotes and **23S** and **5S** in prokaryotes occur in larger subunit of ribosomes while **18S** in eukaryotes and **16S** in prokaryotes is found in smaller subunit of ribosomes.
- **Antisense RNA** also called **micro RNA (*mi*RNA)**, *i.e.*, an RNA inhibiting complementary RNA is synthesized sometimes on the strand complementary to the one used for *m*RNA synthesis. This is used for **regulation of DNA synthesis and gene expression** both *in vitro* and *in vivo*.
- **Small nuclear RNA** (*sn* RNA) is the name used to refer to a number of small RNA molecules found in the nucleus. They are always found associated with specific proteins and the complexes are referred to as small nuclear ribonucleoproteins (*sn* RNPs) or sometime as **snurps**.

DNA versus RNA

- The genetic material should be stable enough not to change with different stages of the life cycle, age or with change in physiology of the organism. DNA being complementary two stranded, if separated by heating come together when appropriate conditions are provided. On the other hand, 2′ **–OH** group present at every nucleotide in RNA is a reactive group makes RNA **labile** and easily **degradable.**
- RNA is also now known to be **catalytic**, hence reactive. Therefore, DNA chemically is less reactive and structurally more stable when compared to RNA. The presence of **thymine** at the place of uracil also confers additional stability to DNA. Both DNA and RNA are able to **mutate.** In fact, RNA being unstable, mutate at the faster rate. Consequently, viruses having RNA genome and having shorter life span mutate and **evolve faster.**

DNA REPLICATION

- The mechanism by which exact copies of the genetic material are formed is known as replication. Replication is necessary so that the genetic informations present in the cells can be carried forward to the daughter cells following cell division.
- The Watson Crick model of DNA suggests that replication of DNA is semiconservative, *i.e.,* the half of the DNA is conserved. **The newly synthesized DNA possesses one strand contributed by parent DNA and other newly synthesized.**
- Semiconservative replication of chromosome was found by **Taylor (1957)** in *Vicia faba* using tritiated thymidine. **Meselson and Stahl (1958)** proved that DNA replicates by semi-conservative method by experimenting on *E.coli* and used the heavy isotopes of nitrogen, *i.e.*, N^{15}. They grew bacterial cells having DNA labelled with ^{15}N in ^{14}N medium and found that F_1 generation has DNA density intermediate between the two.
- DNA replication in prokaryotes and eukaryotes starts from a specific point called **origin of replication** (ori) which is one in bacterial DNA and many in eukaryotic DNA. The region where the helix unwinds and synthesis of new DNA starts is called the **replication fork** and the smallest unit of replication is called **replicon**. Eukaryotic DNA are very large hence they represent several replicon. Bacterial DNA represents only one replicon.
- The replication of DNA requires many enzymes and protein factors which are discussed below:
- **DNA helicase** – DNA helicases are ATP - dependent unwinding enzymes which promote separation of the two parental strands by breaking hydrogen bonds between base pairs and establish replication forks that will progressively move away from the origin.
- **Single strand DNA binding proteins or SSBPs**– Behind the replication fork, the single DNA strand is prevented from rewinding about one another (or forming double stranded hair-pin loops in each single strands) by the action of SSB proteins.
- **Topoisomerases** - Topoisomerase enzymes change the way in which DNA is packaged in living cells by altering the degree of supercoiling of the DNA molecules.

Topoisomerase I removes supercoiling by introducing a temporary break in one DNA strand and passing the other strand through it. Topoisomerase II or DNA gyrase introduces negative supercoiling by breaking both strands of the DNA double helix and passing another stretch of the double helix through the gap, which is then sealed.

- **DNA polymerase or replicase** – DNA polymerase enzymes bring about the synthesis of one polynucleotide chain that is a copy of another. Three different DNA polymerases are known in prokaryotes, of which **DNA polymerase I and II are meant for DNA repair and DNA polymerase III is meant for actual DNA replication.** DNA polymerase I enzyme is also called as **Kornberg enzyme** because it was isolated by **Arthur Kornberg** around 1960. Eukaryotes are found to contain five different types of polymerases, namely, α, β, γ, δ, ε. DNA polymerase γ is also called mitochondrial polymerase.
- **DNA ligases (discovered by H.G. Khorana in 1967)** enzymes capable of catalyzing phosphodiester bond formation between **free 3′ – OH and free 5′ – P** groups of a nick of DNA which is created by endonuclease enzyme, thereby restoring intact DNA duplex.

Mechanism of DNA replication

- During replication the four nucleotides of DNA are activated by ATP in the presence of **phosphorylase**. The existing DNA molecule on which new DNA is synthesized is called as **template DNA**. The first important step in the replication is unwinding of double helix.
- The initiation of DNA synthesis requires a **RNA primer** (a short sequence RNA). The synthesis of RNA primer is brought about by enzyme **primase**. The primer grows in 5′ – 3′ direction. Initiation of replication occurs at 3′ end of the template. The enzyme DNA polymerase adds the nucleotides complementary to the DNA template in 5′ – 3′ direction in the presence of ATP.

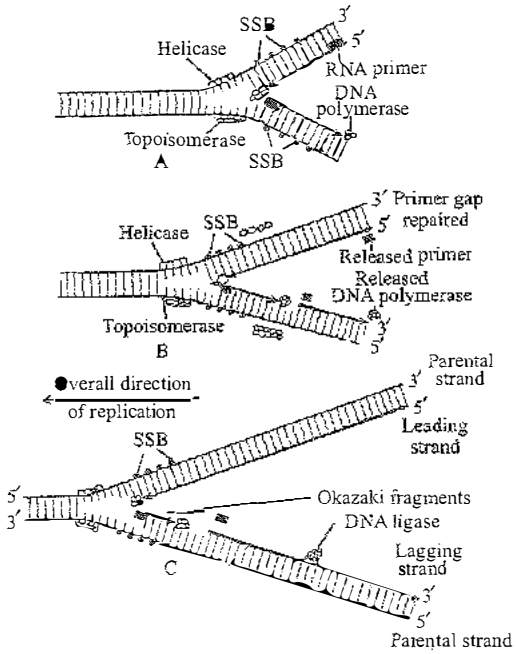


Fig.: Replication of DNA - continuous over one strand and discontinuous over other strand.

- The enzyme synthesizes a new strand in continuous stretch on 3′ – 5′ strand. This strand is called **leading strand**. The second new strand (lagging strand) is formed in short segments called Okazaki fragments on the template strand with polarity 5′– 3′. **Okazaki fragments** are joined by means of **DNA ligase**.
- Such a DNA replication, where the leading strand is synthesized continuously and the lagging strand is synthesized discontinuously, is called **semidiscontinuous replication**. RNA primer is removed by exonuclease activity of **DNA polymerase I**.
- The wrong base entered into DNA helix can be identified and corrected by repair enzymes during proof reading.

GENETIC CODE

- The genetic representation of codon by which the information in RNA is decoded in a polypeptide chain is called genetic code. It is the relationship of amino acid sequences in a polypeptide chain and base sequences of DNA.
- **George Gamow (1954)** was first to propose triplet code and coined the term genetic code.
- **Genetic code was discovered by Nirenberg and Matthaei (1961). Marshell Nirenberg, Severo Ochoa, Har Gobind Khorana, Francis Crick and Matthaei deciphered the genetic code.**
- The 64 distinct triplets determine the 20 amino acids. These triplets are called codons. Any coded message by these condons is commonly called **cryptogram**.

First letter of codon (5′ end)	U		C		A		G		Third letter of codon (3′ end)
U	UUU	Phe	UCU	Ser	UAU	Tyr	UGU	Cys	U
	UUC	Phe	UCC	Ser	UAC	Tyr	UGC	Cys	C
	UUA	Leu	UCA	Ser	UAA	Stop	UGA	Stop	A
	UUG	Leu	UCG	Ser	UAG	Stop	UGG	Trp	G
C	CUU	Leu	CCU	Pro	CAU	His	CGU	Arg	U
	CUC	Leu	CCC	Pro	CAC	His	CGC	Arg	C
	CUA	Leu	CCA	Pro	CAA	Gln	CGA	Arg	A
	CUG	Leu	CCG	Pro	CAG	Gln	CGG	Arg	G
A	AUU	Ile	ACU	Thr	AAU	Asn	AGU	Ser	U
	AUC	Ile	ACC	Thr	AAC	Asn	AGC	Ser	C
	AUA	Ile	ACA	Thr	AAA	Lys	AGA	Arg	A
	AUG	Met	ACG	Thr	AAG	Lys	AGG	Arg	G
G	GUU	Val	GCU	Ala	GAU	Asp	GGU	Gly	U
	GUC	Val	GCC	Ala	GAC	Asp	GGC	Gly	C
	GUA	Val	GCA	Ala	GAA	Glu	GGA	Gly	A
	GUG	Val	GCG	Ala	GAG	Glu	GGG	Gly	G

- Genetic code has following characteristics:
  - **Genetic code is universal** *i.e.*, found in all living organisms.
  - For a particular amino acid more than one codon can be used. This is called **degeneracy of codon**.
  - Genetic codes are **non-overlapping**, *i.e.*, one letter cannot be used for two different codons.
  - A particular codon will always code for same amino acid, *i.e.*, code is **non-ambiguous**. But GUG when present in beginning codes for methionine, but when present in intermediate position, codes for valine.
  - The codon which initiates the protein synthesis is called **initiation codon**. It is mostly AUG which codes for methionine.

- The codon which do not code for any amino acid are called **non-sense codon or terminator codon**. Three codon, *viz*., UAG (amber), UAA (ochre) and UGA (opal) are non-sense codon. In ***Paramecium*** and some other ciliates termination codons UAA and UGA code for glutamine. AGG and AGA code for arginine but function as stop signals in human mitochondrion. UGA, a termination codon, corresponds to tryptophan while AUA denotes methionine in human mitochondria.
- The genetic code is commaless *i.e*., two adjacent codons are continuous.
- The code has polarity Code of *m*RNA is read from 5′ → 3′ direction.

- **Wobble hypothesis** was given by F.H.C. Crick (1965). According to this, third nitrogenous base of a codon is not much significant and codon is specified by first two bases. Hence the same *t*RNA can recognise more than one codons differing only at third position.

CENTRAL DOGMA

- **Crick (1958)** proposed the central dogma of molecular biology. Central dogma is the unidirectional flow of information from DNA to RNA and from RNA to polypeptide.
- **Commoner (1968)** suggested the circular flow of information.
- **H. Temin and Baltimore (1970)** introduced the concept of reverse central dogma, *i.e*., formation of DNA from RNA. It is also called **teminism** and occurs in retroviruses.

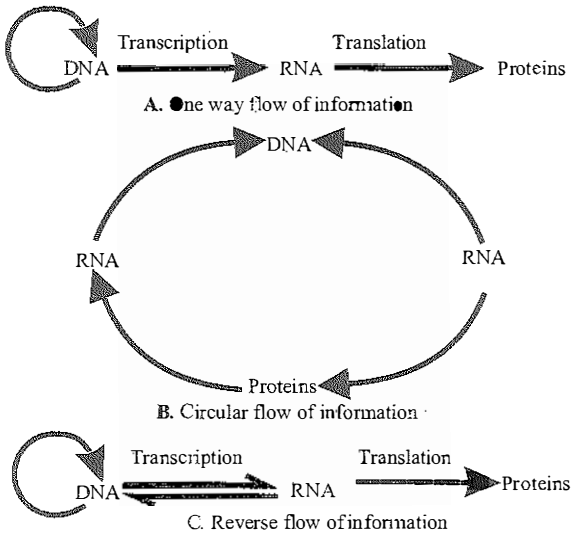


Fig.: Diagrammatic representation of three concepts about the flow of transcriptional information from DNA.

TRANSCRIPTION

- Transcription is the process through which a DNA sequence is enzymatically copied by an RNA polymerase to produce a complementary RNA. Transcription occurs in the nucleus during G_1 and G_2 phases of cell cycle.
- Like DNA replication, transcription proceeds in the 5′ → 3′ direction *i e*., the DNA is read in the 3′ → 5′ direction and the complementary transcript fragments are generated in the 5′ → 3′ direction.
- Transcription requires enzyme RNA polymerase. In prokaryotes single RNA polymerase enzyme undertakes the formation of all RNA. RNA polymerase of *E.coli* has five polypeptides – σ, β, β′ and 2α chains. Chains of β, β′ and 2α chains constitute the core enzyme. The function of σ factor is to confer the **specificity of RNA synthesis** at the promotor site.

  σ or sigma factor recognises the promotor region while the remaining core enzymes take part in transcription.
- In eukaryotes, **three major classes of RNA polymerases** are found in the nucleus. **RNA polymerase I** synthesizes precursors for the large ribosomal RNAs. **RNA polymerase II** synthesizes the precursors for *m*RNAs and small nuclear RNAs. **RNA polymerase III** participates in the formation of *t*RNAs and small ribosomal RNAs.
- In bacteria, transcription takes place in three steps which are **initiation, elongation** and **termination**.
- The RNA polymerase is only capable of catalyzing the process of elongation. It associates transiently with **initiation factor** (σ) and **termination-factor** (ρ) to initiate and terminate the transcription respectively. Association with these factors alter the specificity of the RNA polymerase to either initiate or terminate.
- RNA polymerase binds to promoter and initiates transcription (**initiation**). It also facilitates opening of the helix and continues **elongation**. Once the polymerases reaches the terminator region, the nascent RNA falls off and also the RNA polymerase. This results in **termination** of transcription.
- In bacteria transcription and translation take place in the same compartment and *m*RNA does not require any processing to become active.
- In eukaryotes, there are additional complexities. There are at least **three RNA polymerases** in the nucleus and secondly the primary transcripts contain both the **exons (expressing sequences)** and the **introns (interrupting sequences)** that are the non-functional units. Hence, it is subjected to a process called splicing where the introns are removed and exons are joined in a defined order.
- **hnRNA** (heterogenous nuclear RNA, precursor of *m*RNA) undergoes additional processing called as **capping** and **tailing**. In capping, an unusual nucleotide (methyl guanosine triphosphate) is added to the 5′ – end of hnRNA. In tailing, adenylate residues (200-300) are added at 3′ – end in a template in independent manner. This processing helps in keeping *m*RNA away from degrading action of nucleases present in nucleus and also in easy transportation out of the nucleus.

TRANSLATION

- Translation or protein synthesis is a process during which the genetic information (which is stored in the sequence of nucleotides in an *m*RNA molecule) is translated. It follows the dictations of the genetic code, into the sequence of amino acids in the polypeptide, requiring the functions of a large number of macromolecules.
- Protein synthesis occurs over the **ribosomes**. Each ribosome has two unequal parts, small and large. The **larger subunit** of ribosomes has a groove for pushing out the newly formed polypeptide and protecting the same

from cellular enzymes. The **smaller subunit** fits over the larger one like a cap but leaves a tunnel for *m*RNA.

- The two subunits come together only at the time of protein formation. The phenomenon is called **association**. Mg^{2+} is essential for it. Ribosomes usually form helical groups during active protein synthesis, known as **poiyribosomes** or **polysomes.** Ribosome also acts as a catalyst (23S *r*RNA in bacteria is enzyme ribozyme) for formation of peptide bond.
- It is a messenger RNA which brings **coded information** from DNA and takes part in its translation by bringing amino acids in a particular sequence during the synthesis of polypeptide. The codons of *m*RNA are recognised by anticodons of *t*RNAs.
- The synthesis of polypeptide can be considered in terms of **initiation, elongation,** and **termination stages.** These fundamental processes have additional stages : **activation of amino acids** before their incorporation into polypeptide and the **post-translational processing** of the completed polypeptide. Both these processes play important roles in ensuring the proper function of the protein product.
- For the synthesis of a polypeptide the **carboxyl group** of each amino acid must be activated to facilitate formation of a peptide bond. Amino acids are activated in the **presence of ATP** and linked to their cognate *t*RNA – a process commonly called as charging of *t*RNA or **aminoacylation of *t*RNA** to be more specific.
- The translation of *m*RNA begins with the formation of **initiation complex.** Initiation factors are designated as **IFs in prokaryotes** and eIFs in eukaryotes. Elongation requires the initiation complex, aminoacyl-*t*RNAs, **elongation factors** and GTP.
- During **elongation stage,** complexes composed of an amino acid linked to *t*RNA, sequentially bind to the appropriate codon in *m*RNA by forming complementary base pairs with the *t*RNA anticodon. The ribosome moves from codon to codon along the *m*RNA.
- **Termination** is signaled by the presence of one of three **termination codons** in the *m*RNA **(UAA, UAG, UGA),** immediately following the final coded amino acid. Three release factors **RF-1, RF-2,** and **RF-3** contribute in release of the free polypeptide and last *t*RNA, now uncharged, from the ribosome. Polypeptide formed consists of specific sequence of amino acids. The process in repeated several times creating polyribosomes or polysomes. mRNA has some additional sequences that are not translated are referred to as **untranslated regions** (UTRs), required for efficient translation.
- **Puromycin** acts as an analogue of aminoacyl-*t*RNA. Thus, it is a potent inhibitor of protein synthesis.

REGULATION OF GENE EXPRESSION

- **Gene expression** is the mechanism at the molecular level by which a gene is able to express itself in the phenotype of an organism. **Gene regulation** is the mechanism of switching off and on of the genes depending upon the requirement of the cells and the state of development.
- There are two types of gene regulations **positive** and **negative.** In negative gene regulation, the genes continue to express their effect till their activity is suppressed. This type of gene regulation is also called **repressible regulation.** The repression is due to a product of **regulatory genes.** Positive gene regulation is the one in which the genes remain non-expressed unless they are induced to do it. It is, therefore, **inducible regulation.** Here a product removes a biochemical that keeps the genes in non-expressed state.
- As the genes express their effect through enzymes, their enzymes are also called **inducible enzymes** and **repressible enzymes.**
- Gene regulation is exerted at **four levels** :
  - Transcriptional level when primary transcript is formed.
  - Processing level when splicing and terminal additions are made.
  - Transport of *m*RNA out of nucleus into cytoplasm.
  - Translational level.

Gene regulation in prokaryotes

- The regulation of gene expression in bacteria is carried out through operon system. Operon model was first given by two French microbiologist **Francis Jacob and Jacques Monod (1961)** in *E.coli,* for which they were awarded Nobel Prize in 1965. Operon is a segment of genetic material (DNA) which functions as regulated units *i.e.,* units that can be switched on or switched off.
- An operon system consists of **structural genes, operator gene, regulator gene, promoter gene, a repressor and an inducer or corepressor** (from outside).
- The segment of a DNA molecule determining the amino acid sequences of a protein is known as **structural gene.** Structural genes produce *m*RNA which forms polypeptides/proteins/enzymes under the operational control of an **operator gene.**
- **Operator gene** gives passage to RNA polymerase moving from promoter to structural gene, *i.e.,* it controls the activity of structural genes.
- **Regulator gene** controls activity of operator gene by producing repressor molecules.
- **Promoter gene** is an initiation point for transcription and the site for binding of RNA polymerase.
- Operons can be of two types inducible and repressible.

Inducible operon

- Inducible operon system remains switched off normally but becomes operational in the presence of **inducer,** *e.g.,* *lac* operon system. Inducible operon system occurs in catabolic pathway.
- ***Lac* operon** of *E.coli* has 3 structural genes – **Z, Y and A** which produce 3 enzymes for the degradation of lactose to glucose and galactose. *Z* produces β-galactosidase for splitting lactose into glucose and galactose. Y produces **β-galactoside permease** (membrane bound protein) which is required for the entry of the lactose. A produces **β-galactoside transacetylase** enzyme that transfers an acetyl group from acetyl CoA to β-galactosides.
- In the *lac* operon **allolactose** is the actual inducer while lactose is the apparent inducer.
- *E. coli* contains a protein called the **catabolic activator protein (CAP).** CAP and cAMP bind to one another to

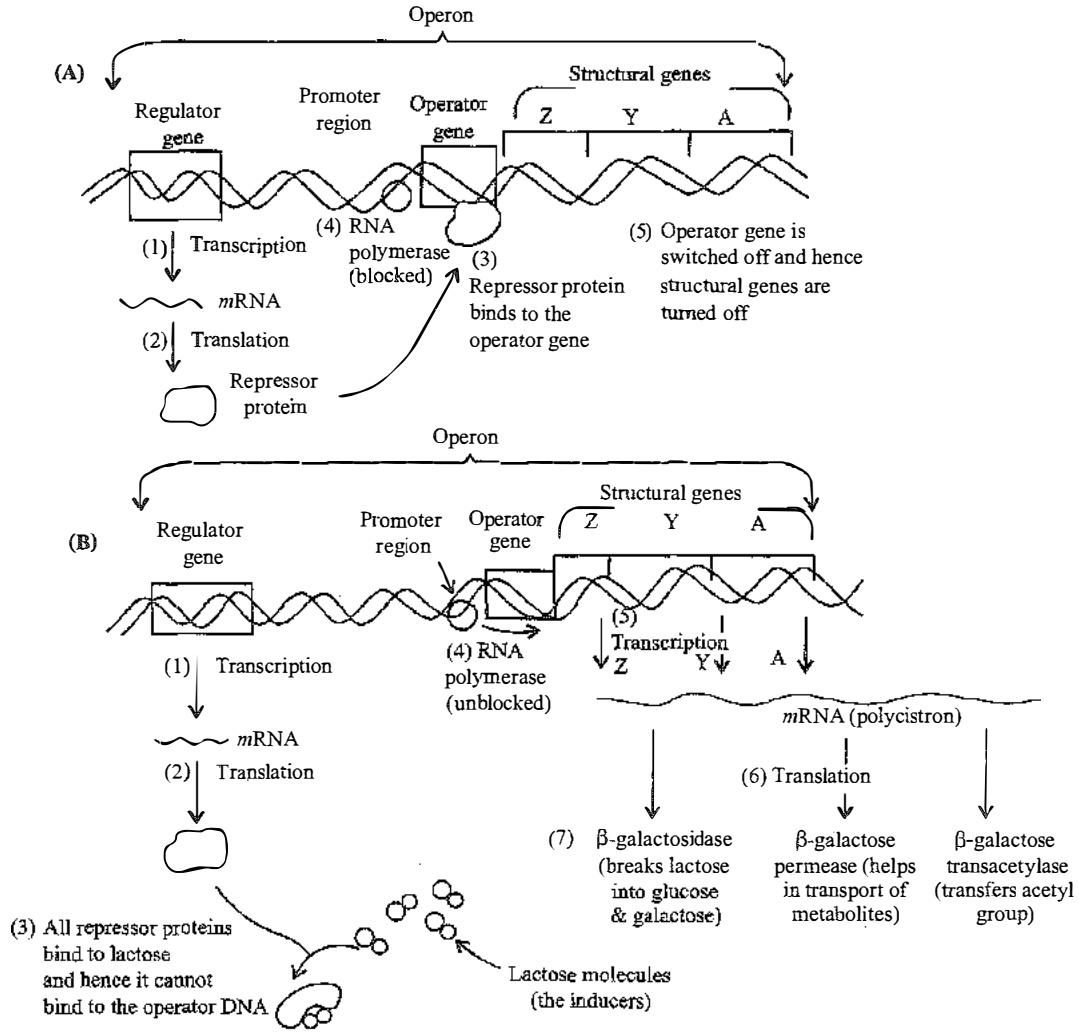


Fig. : Mechanism of *lac* operon (A) Switched off; (B) Switched on

form a unit, called cAMP-CAP which is an active regulatory element of the *lac* system. The **cAMP-CAP complex** must be bound to a base sequence in the DNA in the promotor region in order for transcription to occur. The cyclic AMP (cAMP) are present in *E.coli* when concentration of glucose is quite low in culture media.

Repressible operon

- Repressible operon system normally remains active but can be switched off when the cell does not require the metabolite or the concentration of an end product crosses a threshold value, *e g*, tryptophan operon system. Mechanism of repressible operon is explained as follows :
- **In the absence of corepressor (tryptophan)**

Regulator gene
↓
Inactive repressor
(aporepressor, proteinaceous substance)
↓
No blocking of operator gene
↓
*m*RNA transcription by structural genes
↓
Enzyme synthesis

- **In the presence of corepressor (tryptophan)**

Regulator gene
↓
Inactive repressor (aporepressor)
+
Corepressor (non-protein compound)
↓
Active repressor
↓
Binds to and blocks operator gene
↓
No *m*RNA transcription by structural gene
↓
Five structural genes are not expressed
(hence no enzymes are formed)

Gene regulation in eukaryotes

- In eukaryotes gene regulation takes place by a different mechanism. The possible candidates for regulation of gene activity are the proteins associated with DNA in the chromosomes. These include histones and non-histone chromosomal proteins.

HUMAN GENOME PROJECT (HGP)

- HGP is an undertaking project by six countries (administered jointly by the National Institute of Health and Department of Energy, U.S.A) to acquire "**complete knowledge of the organization, structure, and function of the human genome**". It is called International Human Genome Sequencing Consortium and is aimed at finding out all the genes in each of the human chromosomes (and ultimately of other organisms), determining their function, and hopefully understanding how they together form the complete organism.
- This is regarded as the most ambitious project ever undertaken by humans. It involved over 2,800 scientists working in 20 different institutions in France, Germany, Japan, China, U.K. and U.S.A.
- Discussions among scientists to undertake this project began in the mid 1980s. In 1988, the National Institute of Health, Bethesda, Maryland established an office of Human genome project with James Watson as its first director. The Human genome project officially began on October 1, 1990, and was completed in April, 2003.
- The goals of the genome project are as follows :
  (i) To develop a genetic linkage map of human genome by identifying thousands of genetic markers and mapping them in the genome.
  (ii) To obtain a physical map of human genome by cloning genomic DNA into YACs and cosmids.
  (iii) To sequence the entire human genome.
  (iv) To develop technology for the management of human genome information by developing user friendly tools to provide scientists easy access to up to date information generated by the project, and by developing analytical tools to interpret this information.
  (v) To analyze the genome of other model organisms, *e.g.*, *E. coli*, *Drosphila melanogaster*, *Arabidopsis thaliana*, *Mus musculus* (mouse), etc.
  (vi) To develop programmes focussed on understanding and addressing the ethical, legal, and social implications of the results obtained from the project.
  (vii) To develop technological advances in genetic methodologies, *e.g.*, gene cloning, DNA sequencing etc.

Salient features of human genome

(i) Human genome has 3.1647 billion nucleotide base pairs.
(ii) The average gene size is 3000 base pairs. The largest gene is that of Duchenne Muscular Dystrophy on X-chromosome. It has 2.4 million (2400 kilo) base pairs. β-globin and insulin genes are less than 10 kilobases.
(iii) The human genome consists of about 30,000 genes. The size of genome or number of genes is unconnected with the complexity of body organisation *e.g.*, Lily has 18 times more DNA than human genome, yet it produces fewer protein than us because its DNA has more introns and less exons.
(iv) Chromosome I has 2968 genes while Y-chromosome has 231 genes. They are the maximum and minimum genes for the human chromosomes.
(v) Less than 2% of the genome represents structural genes that code for proteins.
(vi) 99.9% of the nucleotide bases are exactly similar in all human beings.
(vii) Scientists have identified about 1.4 million locations where single base DNA differences (SNPs- single nucleotide polymorphism) occur in humans. This information promises to revolutionize the processes of finding chromosomal locations for disease associated sequences and tracing human history.

DNA FINGERPRINTING OR GENETIC FINGERPRINTING

- DNA fingerprinting is a technique for identifying repeated sequences in the human genome that produce a **pattern of bands unique for every individual**. Science of DNA fingerprinting was first used by **Sir William Herschel** as a method of identification in 1858. Repetitive DNA are part of DNA containing the same sequence of nitrogen bases repeated several times in tandem.
- Satellite DNA are part of repetitive DNA which has long repetitive sequences in tandem that forms a separate fraction on density gradient centrifugation. Depending upon the number of base pairs involved in repeat regions, satellite DNA is of two types, **microsatellite sequences** and **minisatellite sequences**.
- Satellite DNA shows very high degree of polymorphism. It is called as **variable number of tandem repeats** (VNTR). Polymorphism (variation at genetic level) arises due to mutations. In DNA polymorphism more then one variant (allele) at a locus occurs in human population with a frequency greater than 0.01. In simple terms, if an inheritable mutation is observed in a population at high frequency, it is referred to as **DNA polymorphism**.

Methodologies used for genome sequencing

- Expressed sequence tags (ESTs) – Approach focusses on identifying all the genes that are expressed as RNA.
- Sequence annotation – Total DNA from cell is isolated, converted into fragments cloned in suitable host using special vectors. Cloning results amplification and is used for sequencing the genome.
- DNA fingerprints can be prepared from extremely minute amounts of blood, semen, hair bulb or any other cells of the body. The major steps are as follows :
  – DNA is extracted from the cells. DNA can be amplified by **PCR or polymerase chain reaction**.
  – DNA is cut into fragments with **restriction enzymes**.
  – Chopped DNA fragments are passed through **electrophoresis**. The separated fragments can be visualised by staining them with a dye.
  – Double-stranded DNA is then split into single-stranded DNA using **alkaline chemicals**.
  – Separated DNA sequences are transferred from gel onto a **nitrocellulose or nylon membrane** (Southern blotting).
  – The nylon sheet is then immersed in a bath, where probes or markers (radioactive synthetic DNA segments of known sequences) are added. The probe hybridises VNTR.

- X-ray film is exposed to the nylon sheet which gives dark bands at the probe sites. Thus hybridised fragments are detected by autoradiography.

- DNA fingerprinting helps in:
  - identifying the true (biological) father.
  - identification of the criminal in rape, murder cases etc.
  - solving the problems of evolution.

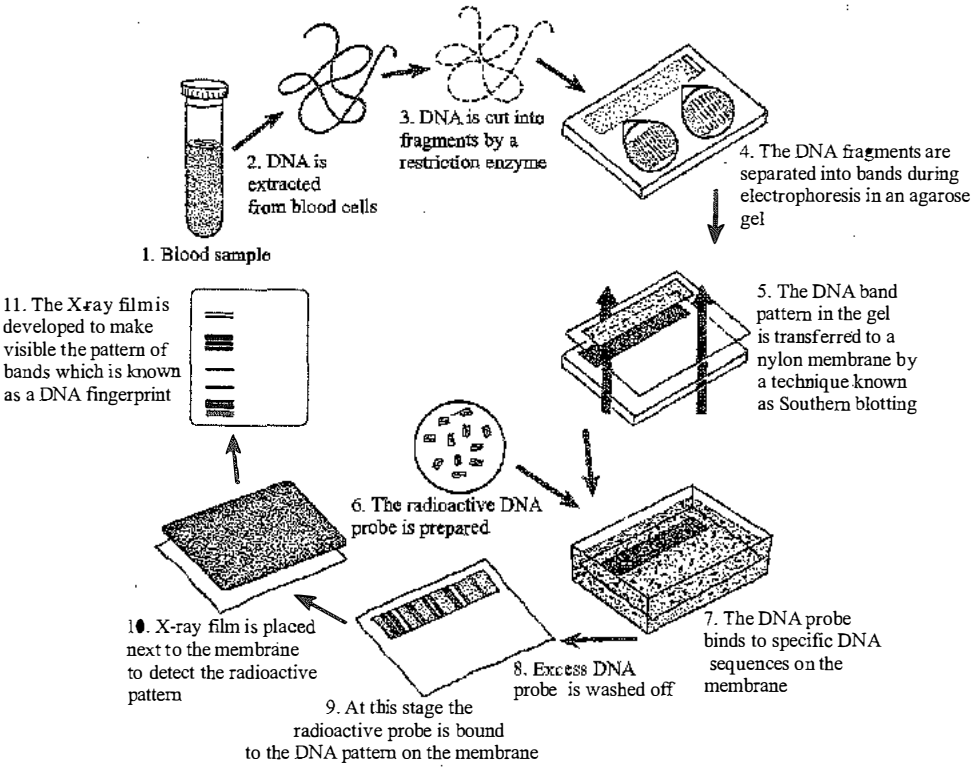


Fig.: The DNA fingerprinting process.

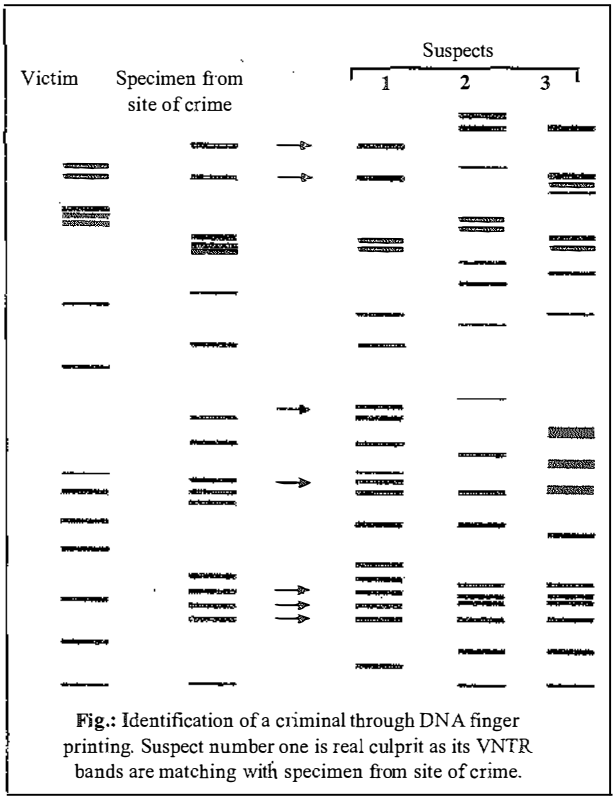


Fig.: Identification of a criminal through DNA finger printing. Suspect number one is real culprit as its VNTR bands are matching with specimen from site of crime.

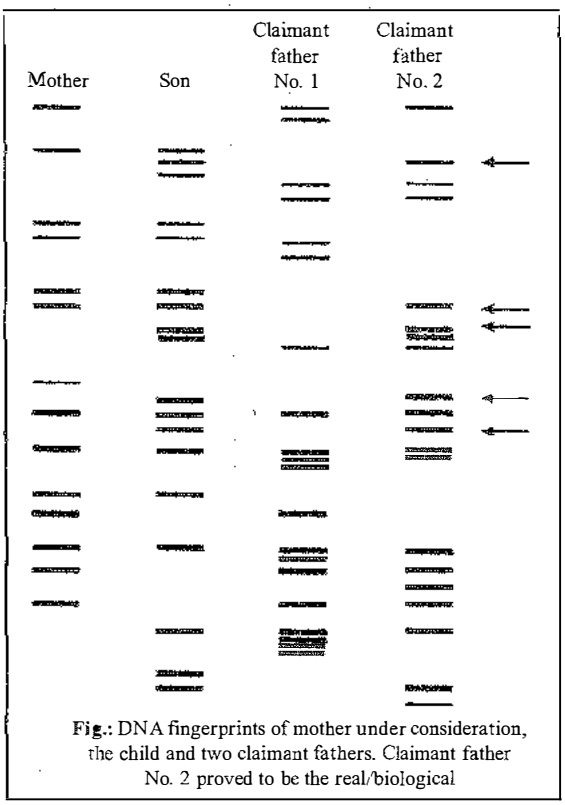
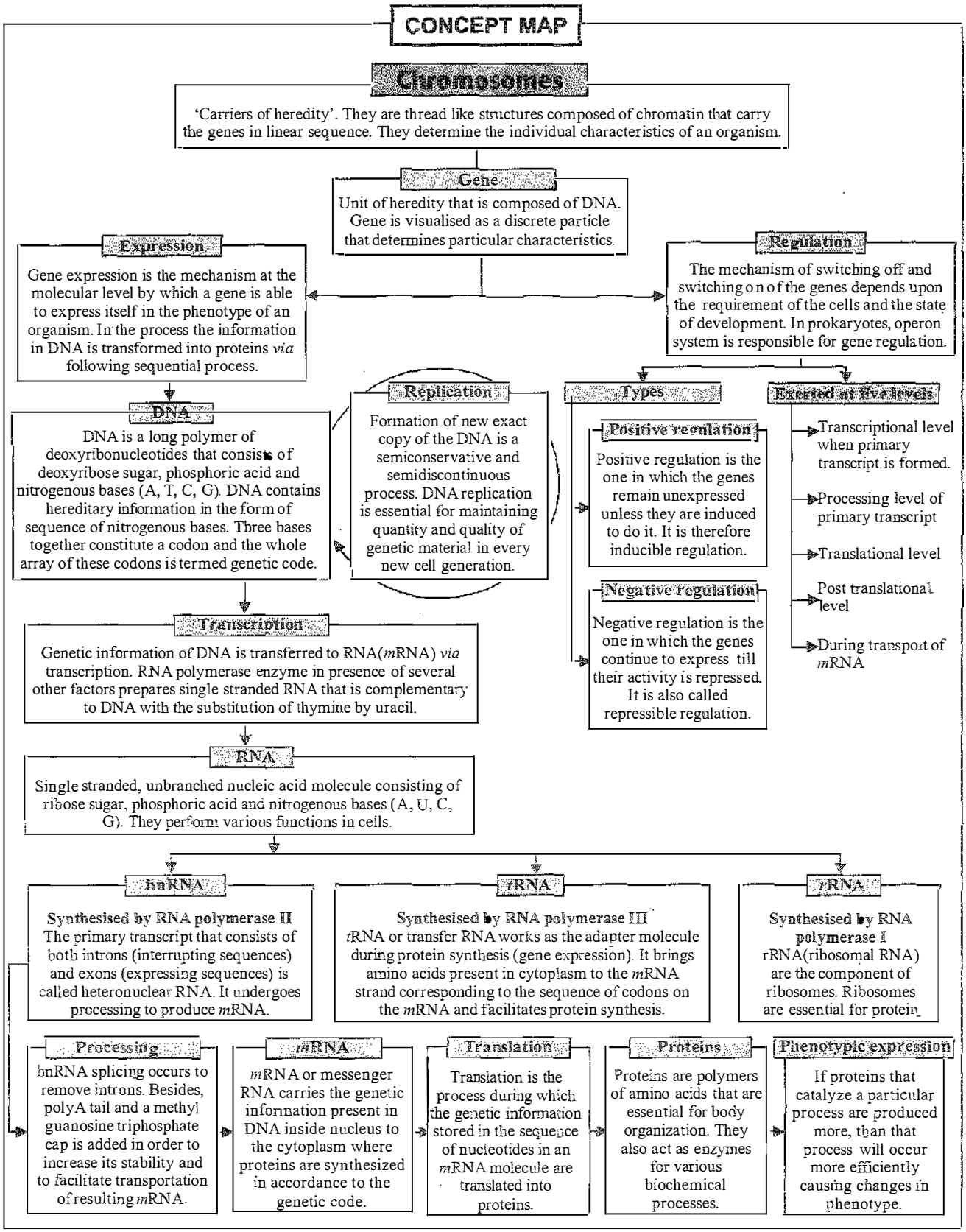


Fig.: DNA fingerprints of mother under consideration, the child and two claimant fathers. Claimant father No. 2 proved to be the real/biological


CONCEPT MAP
Chromosomes
'Carriers of heredity'. They are thread like structures composed of chromatin that carry the genes in linear sequence. They determine the individual characteristics of an organism.
Gene
Unit of heredity that is composed of DNA. Gene is visualised as a discrete particle that determines particular characteristics.
Expression
Gene expression is the mechanism at the molecular level by which a gene is able to express itself in the phenotype of an organism. In the process the information in DNA is transformed into proteins via following sequential process.
Regulation
The mechanism of switching off and switching on of the genes depends upon the requirement of the cells and the state of development. In prokaryotes, operon system is responsible for gene regulation.
DNA
DNA is a long polymer of deoxyribonucleotides that consists of deoxyribose sugar, phosphoric acid and nitrogenous bases (A, T, C, G). DNA contains hereditary information in the form of sequence of nitrogenous bases. Three bases together constitute a codon and the whole array of these codons is termed genetic code.
Replication
Formation of new exact copy of the DNA is a semiconservative and semidiscontinuous process. DNA replication is essential for maintaining quantity and quality of genetic material in every new cell generation.
Types
Positive regulation
Positive regulation is the one in which the genes remain unexpressed unless they are induced to do it. It is therefore inducible regulation.
Negative regulation
Negative regulation is the one in which the genes continue to express till their activity is repressed. It is also called repressible regulation.
Exerted at five levels
Transcriptional level when primary transcript is formed.
Processing level of primary transcript
Translational level
Post translational level
During transport of mRNA
Transcription
Genetic information of DNA is transferred to RNA(mRNA) via transcription. RNA polymerase enzyme in presence of several other factors prepares single stranded RNA that is complementary to DNA with the substitution of thymine by uracil.
RNA
Single stranded, unbranched nucleic acid molecule consisting of ribose sugar, phosphoric acid and nitrogenous bases (A, U, C, G). They perform various functions in cells.
hnRNA
Synthesised by RNA polymerase II
The primary transcript that consists of both introns (interrupting sequences) and exons (expressing sequences) is called heteronuclear RNA. It undergoes processing to produce mRNA.
tRNA
Synthesised by RNA polymerase III
tRNA or transfer RNA works as the adapter molecule during protein synthesis (gene expression). It brings amino acids present in cytoplasm to the mRNA strand corresponding to the sequence of codons on the mRNA and facilitates protein synthesis.
rRNA
Synthesised by RNA polymerase I
rRNA(ribosomal RNA) are the component of ribosomes. Ribosomes are essential for protein
Processing
hnRNA splicing occurs to remove introns. Besides, polyA tail and a methyl guanosine triphosphate cap is added in order to increase its stability and to facilitate transportation of resulting mRNA.
mRNA
mRNA or messenger RNA carries the genetic information present in DNA inside nucleus to the cytoplasm where proteins are synthesized in accordance to the genetic code.
Translation
Translation is the process during which the genetic information stored in the sequence of nucleotides in an mRNA molecule are translated into proteins.
Proteins
Proteins are polymers of amino acids that are essential for body organization. They also act as enzymes for various biochemical processes.
Phenotypic expression
If proteins that catalyze a particular process are produced more, than that process will occur more efficiently causing changes in phenotype.