Module 1

Bioinformatics and Computational Biology, Nature & Scope of Bioinformatics. The central dogma of molecular biology and bio-sequences associated with it, RNA classification –coding and non coding RNA-mRNA, tRNA, miRNA and sRNA, RNAi. DNA and RNA structure – Nucleic Acid structure and function, Genetic Code, Genes and Evolution.

Computational Biology and Bioinformatics

Computational biology and bioinformatics is an interdisciplinary field that develops and applies computational methods to analyse large collections of biological data, such as genetic sequences, cell populations or protein samples, to make new predictions or discover new biology. The computational methods used include analytical methods, mathematical modelling and simulation.

Bioinformatics: Research, development, or application of computational tools and approaches for expanding the use of biological, medical, behavioral or health data, including those to acquire, store, organize, archive, analyze, or visualize such data.

Computational Biology: The development and application of data-analytical and theoretical methods, mathematical modeling and computational simulation techniques to the study of biological, behavioral, and social systems.

Definition bioinformatics:

Bioinformatics is conceptualizing biology in terms of molecules(in sense from chemistry) and applying informatics techniques(derived from mathematics and computer science)to understand and organize the information associated with this molecules.

Nature and scope of bioinformatics

Bioinformatics canbe used for a wide variety of applications

Basically bioinformatics is used to organize biological data to help the researchers access information, add new information arising out of experiments and modify existing information. The bioinformatics analysis is applicable to genome sequences, macromolecular structure, data from functional genomics experiments, phylogenetic trees, metabolic pathways, the text of scientific papers and medical information. Second level of bioinformatics applications is to develop tools. Third level of bioinformatics is to use these tools to analysis the data to interpret the results in a biologically meaningful manner. Various fields using bioinformatics include molecular medicine, personalized medicine, preventive medicine, gene therapy, drug development, microbial genome applications, climate change studies, alternative energy resources, biotechnology, antibiotic resistance, forensic analysis of microbes, evolutionary studies, crop improvement, insect resistance etc.(Lalitha, 2005). Some of the bioinformatics applications (Rastogi *et al.*, 2004) are discusses in the following sections.

i) Information search and retrieval

It is one of the most powerful applications of bioinformatics. For e
(http://www.ncbi.nlm.nih.gov/PubMed/) is a service of the National Libra
provides access to over 12 million MEDLINE citations back to the mid 196
life science journals. Pub Med includes links to many sites providing full
other related resources.

Etblast (http://chaos.swmed.edu/etblast/index.shtml) is an applicati
compare a query set of sentences with a database of other text to identi
database that is most similar to the query.

Entrez (http://www.ncbi.nlm.nih.gov/Entrez/) is a widely used interfa

ii) Sequence comparison

It is one of the most popular applications for the researchers. The mos
sequence comparison or similarity search are BLAST and FASTA. These
pair-wise comparison of sequences.

iii) Sequence alignment

Sequence alignment software include:

- FASTA (pairwise alignment)
- BLAST (pairwise alignment)

  i) BLASTP- queries a protein sequence against a protein database

iii) BLASTX- queries a translated nucleotide sequence against a pro
iv) TBLASTN- queries a protein sequence against a translated nucle
v) TBLASTX- queries a translated nucleotide sequence against a trans
database

- CLUSTAL W (multiple alignment)
- MULTAL (multiple alignment)

iv) Linkage analysis

Linkage analysis is used to identify the chromosomal location of g
several tools available for linkage analysis and many of these are
linkage.rockefeller.edu/

v) Phylogenetic analysis

It is also known as molecular taxonomy. It uses the representation

vi) Genomics

It refers to mapping, sequencing and analysis of genomes. Structural ge
linkage analysis, molecular cytogenetics, physical mapping and genome sequen
genomics includes gene expression, forward genetics, reverse genomic
genomics and proteomics.

vii) Proteomics

It refers to the study of proteomes (i.e. all proteins produced by a speci
with time.

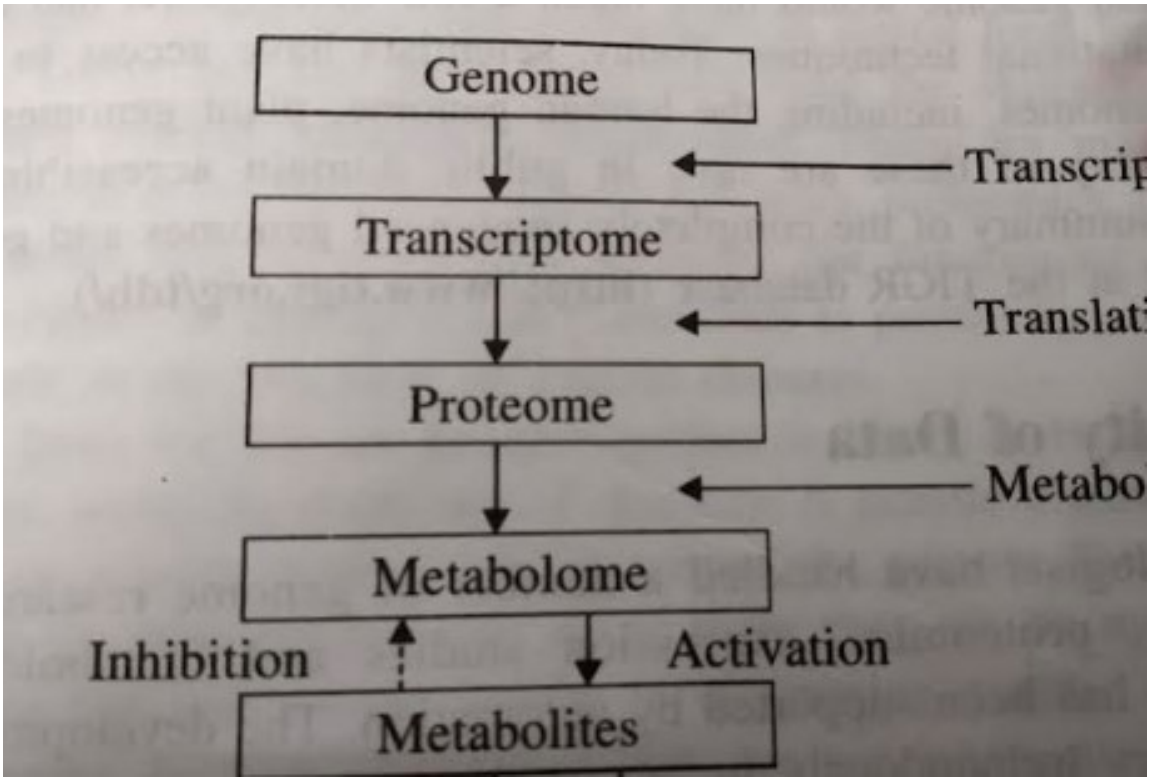
viii) Drug discovery and computer aided drug design

Drug discovery process usually starts with an analysis of binary sites in
or an identification of structural features common to active compounds. T

All we know cells are the basic unit of construction of our human bodyDuring the 1950s, scientists developed the concept that all organisms may be classified as **prokaryotes** or **eukaryotes.** The cells of all prokaryotes and eukaryotes possess two basic features: a plasma membrane, also called a cell membrane, and cytoplasm. However, the cells of prokaryotes are simpler than those of eukaryotes. For example, prokaryotic cells lack a nucleus, while eukaryotic cells have a nucleus. Prokaryotic cells lack internal cellular bodies (organelles), while eukaryotic cells possess them. Examples

of prokaryotes are bacteria and **archaea.** Examples of eukaryotes are protists, fungi, plants, and animals (everything except prokaryotes).

Levels of biological activity in a cell



Genome

A genome is an organism's complete set of DNA, including all of its genes. Each genome contains all of the information needed to build and maintain that organism.It is thw complete set of DNA including genes

Transcriptome

The **transcriptome** is the set of all RNA molecules in one cell or a population of cells. It is sometimes used to refer to all RNAs, or just mRNA, depending on the particular experiment.

Proteome

The proteome is the full complement of proteins produced by a particular genome.

Metabolome

A metabolome is all of the metabolites produced by a single organism. Metabolites are substances required for, or produced by, the biochemical reactions of metabolism in living organisms.

A **metabolites**

is the intermediate end product of metabolism. The term *metabolite* is usually restricted to small molecules. Metabolites have various functions, such as ..including fuel, structure, signaling, stimulatory and inhibitory effects on enzymes, catalytic activity of their own (usually as a cofactor to an enzyme), defense, and interactions with other organisms (e.g. pigments, odorants, and pheromones).

Central Dogma of molecular biology

The genetic material is stored in the form of DNA in most organisms. In humans, the nucleus of each cell contains 3×10^9 base pairs of DNA distributed over 23 pairs of chromosomes, and each cell has two copies of the genetic material. This is known collectively as the human genome

The *Central Dogma of Molecular Biology* states that **DNA makes RNA makes proteins** (Figure 1).

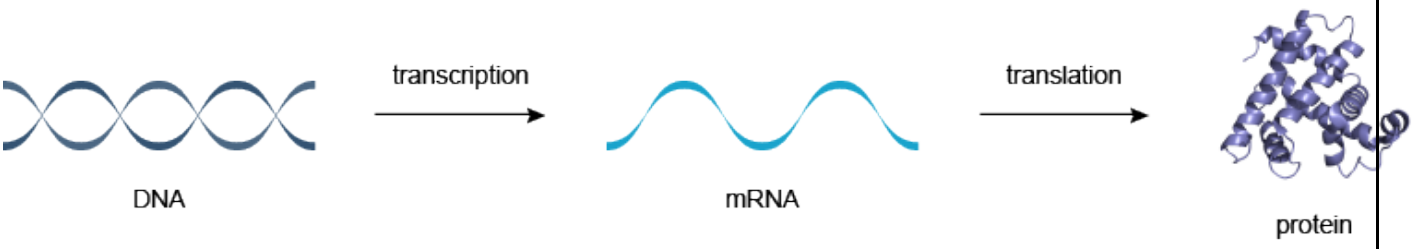


Figure 1 | The Central Dogma of Molecular Biology: DNA makes RNA makes proteins

Transcription

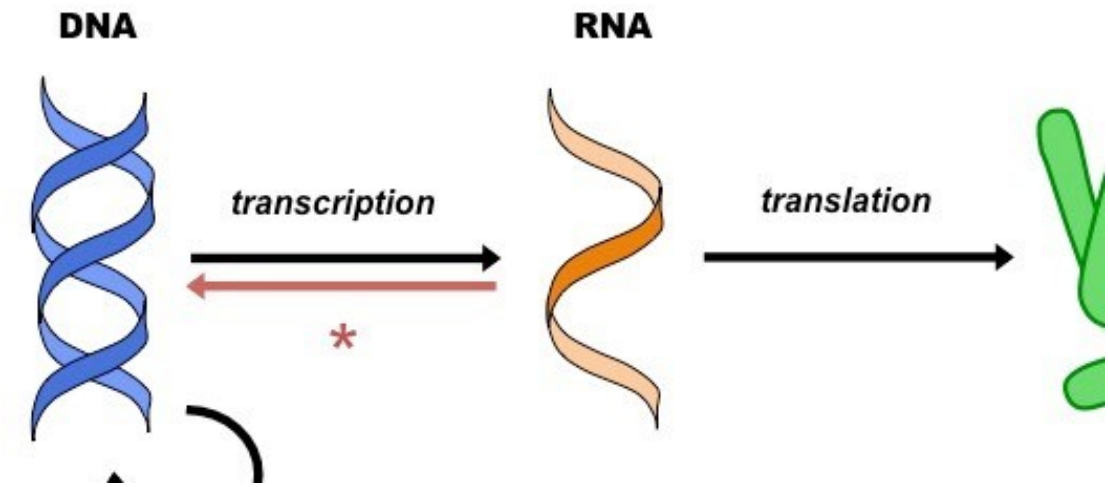
The process by which DNA is copied to RNA is called transcription

Translation

RNA is used to produce proteins by a process called translation.

REPLICATION

The process of creating multiple copies of DNA itself during cell division is called replication



This information flow was always considered to be uni-directional: DNA → RNA → Protein

However in 1970 it was discovered that retroviruses could copy DNA from an RNA sequence

- These viruses possess an enzyme (*reverse transcriptase*) that allows for reverse transcription to occur
- Reverse transcription is now commonly used in scientific studies to establish gene expression profiles

Bio-sequences Associated with molecular biology

Nucleic acids

Nucleic acids are the main information-carrying molecules of the cell, and, by directing the process of protein synthesis, they determine the inherited characteristics of every living thing. The two main classes of nucleic acids are deoxyribonucleic acid (DNA) and ribonucleic acid (RNA).

DNA

The basic characteristics of living organism is determined by DNA that is produced and passed on to next generation.DNA contains instructions to make proteins. It is a blue print of information composed of a linear array of nucleotides each of which has a base plus a deoxyribose sugar and a phosphate .There are 4 types of bases they

are adenine (A), guanine (G), cytosine (C) and , thymine (T).DNA is in form of double helix .

RNA

RNA is single stranded .It contains ribose instead of deoxyribose.In its sugar phosphate backbone Uracil(U) is found instead of thymine(T). It is the first intermediate in converting the information from DNA into proteins essential for the working of a cell. Some RNAs also serve direct roles in cellular metabolism. RNA is made by copying the base sequence of a section of double-stranded DNA, called a gene, into a piece of single-stranded nucleic acid. This process, called transcription is catalyzed by an enzyme called RNA polymerase.

DNA replication

In molecular biology, **DNA replication** is the biological process of producing two identical replicas of DNA from one original DNA molecule. This process occurs in all living organisms and is the basis for biological inheritance. The cell possesses the distinctive property of division, which makes replication of DNA essential.

DNA is made up of a double helix of two complementary strands. During replication, these strands are separated. Each strand of the original DNA molecule then serves as a template for the production of its counterpart, a process referred to as semi conservative. As a result of semi-conservative replication, the new helix will be composed of an original DNA strand as well as a newly synthesized strand.

Genes

A gene is the basic physical and functional unit of heredity. Genes are made up of DNA. Some genes act as instructions to make molecules called proteins. However, many genes do not code for proteins. In humans, genes vary in size from a few hundred DNA bases to more than 2 million bases. The Human Genome Project estimated that humans have between 20,000 and 25,000 genes.

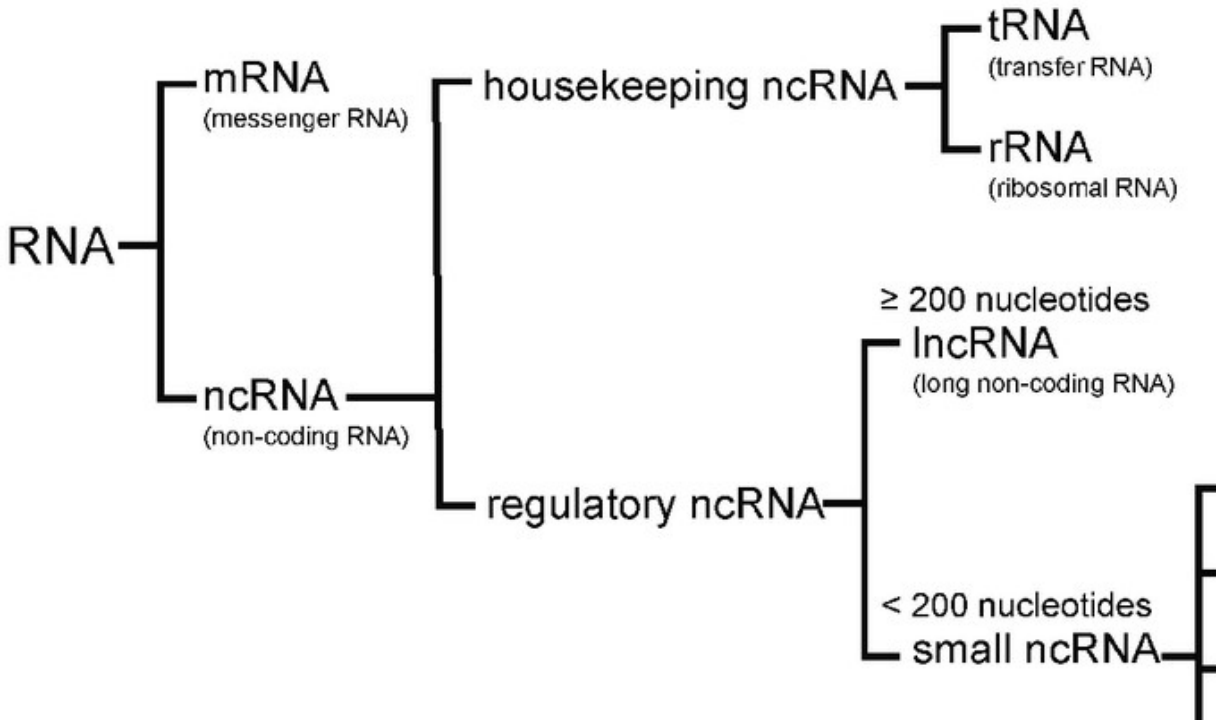
Every person has two copies of each gene, one inherited from each parent. Most genes are the same in all people, but a small number of genes (less than 1 percent of the total) are slightly different between people. Alleles are forms of the same gene with small differences in their sequence of DNA bases. These small differences contribute to each person's unique physical features.

Genome

Genomes constitutes the total genetic material of an organism.Human genome s highly complex and contain about 3 billion nucleotides.

Chromatin

Chromatin is a complex of DNA and proteins that forms chromosomes within the nucleus of eukaryotic cells. Nuclear DNA does not appear in free linear strands; it is highly condensed and wrapped around nuclear proteins in order to fit inside the nucleus.The proteins include positively charged basic proteins called histones as well as non histonic proteins.



Coding RNA

Those RNAs carrying the code for protein synthesis are called “coding RNAs” Surprisingly, recent evidence revealed that very little of our human genome sequences (less than 2%) could actually end up producing proteins.

Messenger RNA(mRNA)

- **Messenger RNA (mRNA)** is the RNA transcript or RNA copy of the DNA message produced during DNA transcription. Messenger RNA is translated to form proteins.

- Messenger RNA (mRNA) carries the genetic code from DNA in a form that can be recognized to make proteins. The coding sequence of the mRNA determines the amino acid sequence in the protein produced. Once transcribed from DNA, eukaryotic mRNA briefly exists in a form called “precursor mRNA (pre-mRNA)” before it is fully processed into mature mRNA. This processing step, which is called “RNA splicing”, removes the introns—non-coding sections of the pre-mRNA. There are approximately 23,000 mRNAs encoded in human genome.

Non-coding RNA (ncRNA)

The Non coding RNA(ncRNAs) do not undergo translation to synthesize proteins, but may hold the key to broaden our understanding of gene regulation and human diseases. Many of them are reported to serve as various regulatory elements in the genome, whereas most are still of unknown importance to gene regulation.

We would like to briefly introduce several different types of non coding RNAs:

Ribosomal RNA (rRNA):

- **Ribosomal RNA (rRNA**) is a component of ribosomes and is also involved in protein synthesis

- Ribosomal RNA is the catalytic component of the ribosomes. In the cytoplasm, rRNAs and protein components combine to form a nucleoprotein complex called the ribosome which binds mRNA and synthesizes proteins (also called translation).

- rRNA, also termed as Ribosomal ribonucleic acid, is the RNA component of the ribosome. It forms an association with protein to form ribosome. It represents seventy percent of cellular RNA and is found in ribosome and nucleolus. rRNA has two types of ribosomes, the Large subunit (LSU) and the small subunit (SSU). It composes 60% of ribosome by weight and plays a critical role in identifying conserved fractions of mRNA and tRNA. It is important for protein synthesis in all the living organisms. rRNA are of ancient origin and can be found in all forms of life.

Transfer RNA (tRNA):

- **Transfer RNA (tRNA)** has a three dimensional shape and is necessary for the translation of mRNA in protein synthesis.

- Transfer RNA is a small RNA chain of about 80 nucleotides. During translation, tRNA transfers specific amino acids that correspond to the mRNA sequence into the growing polypeptide chain at the ribosome.

- Transfer ribonucleic acid (tRNA) is a type of RNA molecule that helps decode a messenger RNA (mRNA) sequence into a protein. tRNAs function at specific sites in the ribosome during translation, which is a process that synthesizes a protein from an mRNA molecule. Proteins are built from smaller units called amino acids, which are specified by three-nucleotide mRNA sequences called codons. Each codon represents a particular amino acid, and each codon is recognized by a specific tRNA. The tRNA molecule has a distinctive folded structure with three hairpin loops that form the shape of a three-leafed clover. One of these hairpin loops contains a sequence called the anticodon, which can recognize and decode an mRNA codon. Each tRNA has its corresponding amino acid attached to its end. When a tRNA recognizes and binds to its corresponding codon in the ribosome, the tRNA transfers the appropriate amino acid to the end of the growing amino acid chain. Then the tRNAs and ribosome continue to decode the mRNA molecule until the entire sequence is translated into a protein.

- The job of tRNA is to decode or translate the language of RNA into the amino acid sequence of proteins. In other words, tRNA is a type of RNA molecule that acts as the carrier of amino acids to the ribosome

Micro RNA (miRNA)

- **MicroRNAs (miRNAs)** are small RNAs that help to regulate gene expression.

- MicroRNAs (miRNAs) are a class of small, endogenous RNAs of 21–25 nucleotides (nts) in length. They play an important regulatory role in animals and plants by targeting specific mRNAs for degradation or translation repression. Recent scientific advances have revealed the synthesis pathways and the regulatory mechanisms of miRNAs in animals and plants. miRNA-based regulation is implicated in disease etiology and has been studied for treatment.

sRNA

- **Bacterial small RNAs** (**sRNA**) are small RNAs produced by bacteria; they are 50- to 500-nucleotide non-coding RNA molecules, highly structured and containing several stem-loops.Numerous sRNAs have

been identified using both computational analysis and laboratory-based techniques

RNAi

RNA interference (RNAi), regulatory system occurring within eukaryotic cells (cells with a clearly defined nucleus) that controls the activity of genes. RNAi functions specifically to silence, or deactivate, genes.

NUCLEIC ACIDS

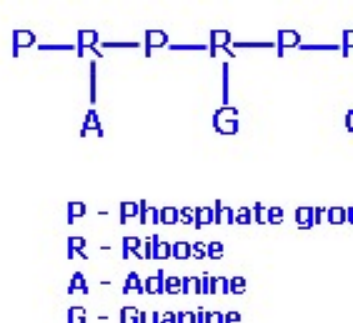
Nucleic acids are involved in the storage and transfer of genetic information in all living organisms, including the simplest viruses. There are 2 types of nucleic acid in cells, ***Deoxyribonucleic acid*** **(DNA)** and ***Ribonucleic acid*** **(RNA)**. Nucleic acids are so named because DNA was 1st isolated from nuclei, but both DNA and RNA also occur in other parts of the cell, e.g. DNA is also found in mitochondria and chloroplasts, whilst RNA is also found in the cytoplasm, particularly at the ribosomes.

Structure of Nucleic acids • NA structure is often divided into four different levels: 1. Primary structure 2. Secondary structure 3. Tertiary structure 4. Quaternary structure

Primary structure: consists of a linear sequence of nucleotides that are linked together by phosphodiester bond.

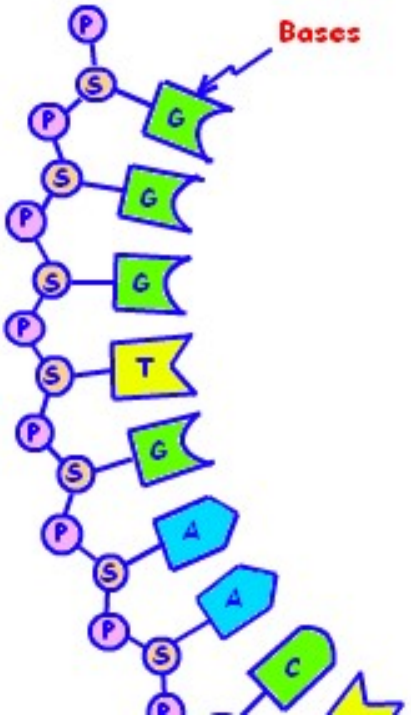
• Nucleotides consists of 3 components:
– Nitrogenous base
– 5-carbon sugar
– One or more phosphate groups

ucleic acid can be identified in terms of different levels of structure. The primary structure of a nucleic acid refers to the sequence of its nucleotide residues. The sugar-phosphate backbone of a given nucleic acid does not vary, the primary structure of the nucleic acid depends only on the sequence of bases present.



Primary structure is determined by the specific content of **adenine (A), thymine (T), guanine (G) and cytosine (C).** Since in template polymerization A and T as well as G and C are mutually complementary, the contents of the complements involved are equal.

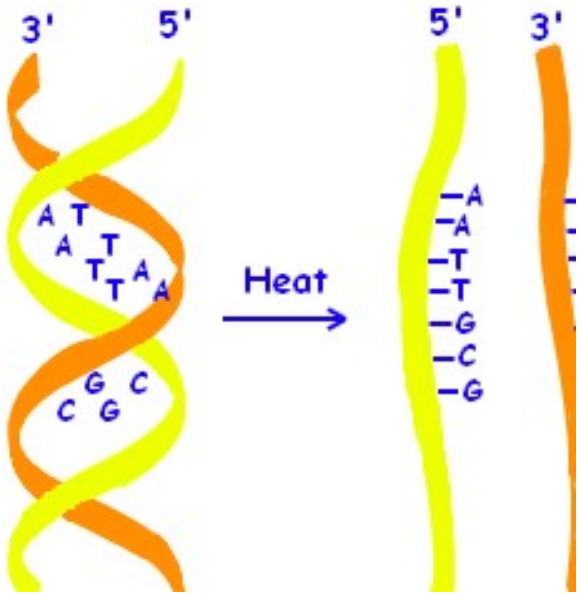
are equal.



Secondary Structure of Nucleic Acid

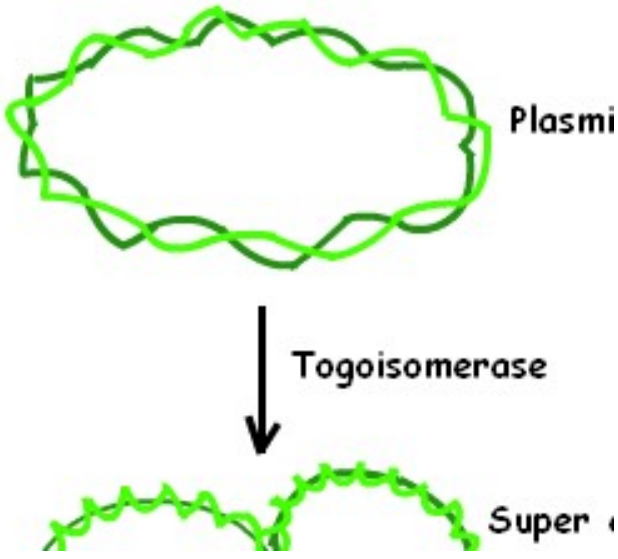
Secondary structure refers to the **α-helix and β-sheet**. Secondary structure arises from the interaction of Van Der Walls forces and in particular of hydrogen bonds.

These form between the oxygen and nitrogen atoms of the complementary pyrimidine and purine bases.
Watson and Crick remains the classic example of secondary structure of nucleic acids.



Tertiary Structure of Nucleic Acid

For nucleic acid tertiary refers to overall three-dimensional shape, including the contribution of secondary structure. In DNA, tertiary structures arises from super coiling, which involves double believes being twisted into tighter more compact shapes



A relatively simple example of super coiling in DNA is that which takes place with plasmids, aparticular form of circular DNA found in bacteria.

Quaternary structure of Nucleic Acid

Nucleic acid quaternary structure refers to the interactions between separate nucleic acid molecules, or between nucleic acid molecules and proteins.

main functions of nucleic acid

- Store and transfer genetic information
- To use the genetic information to direct the synthesis of new protein.
- The deoxyribonucleic acid is the storage for place for genetic information in the cell.
- DNA controls the synthesis of RNA in the cell.

Both DNA and RNA are polymers, the monomeric units being called ***nucleotides.*** DNA and RNA are therefore ***polynucleotides.***

THERE ARE **FIVE** DIFFERENT NITROGEN-CONTAINING ORGANIC BASES PRESENT IN NUCLEIC ACIDS:

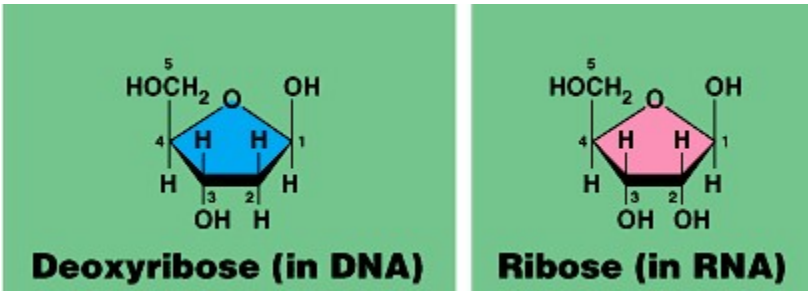
- ***adenine*** (A), ***guanine*** (G), ***cytosine*** (C), ***thymine*** (T) and ***uracil*** (U).
- A, **T**, C and G are found in DNA.
- A, **U**, C and G are found in RNA (uracil replaces thymine here).

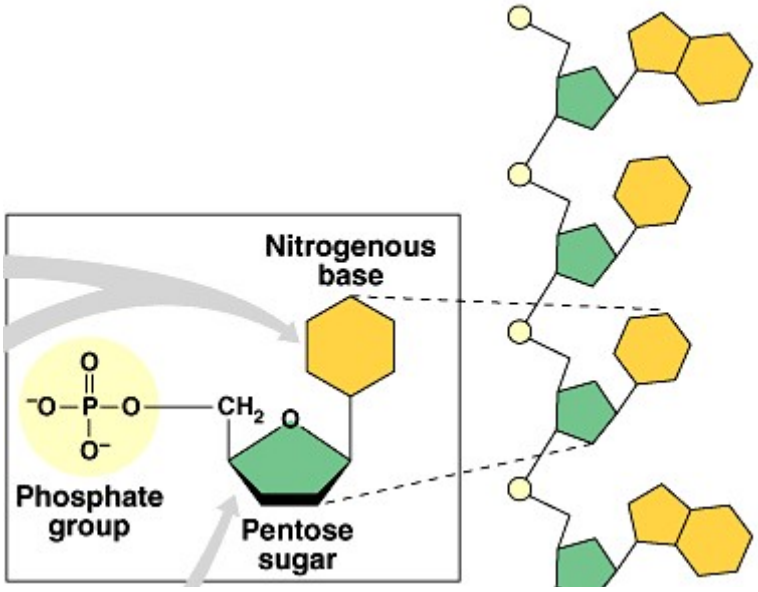
Nucleic acids were made up of Simple units called nucleotides, connected in long chains

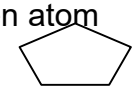
•Nucleotides have 3 parts:

1. 5-Carbon sugar (pentose sugar)
   Deoxyribose in DNA
   Ribose in RNA
2. Nitrogen containing base (made of C, H and N)
   A,U,C,T and U
3. A phosphate group (PO4)

•The P groups make the links that unite the sugars (hence a “sugar-phosphate backbone”)

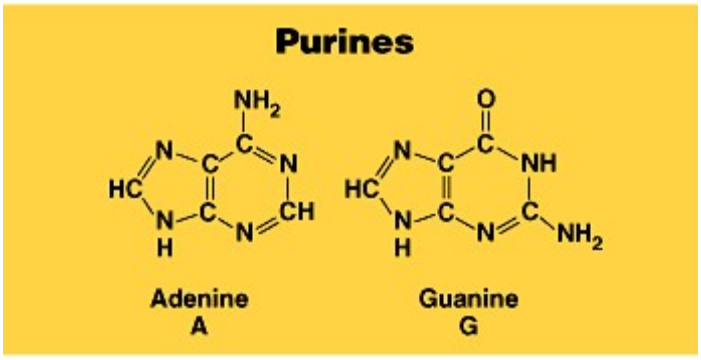




- Ribose is a sugar, like glucose, but with only five carbon atoms in its molecule
- Deoxyribose is almost the same but lacks one oxygen atom
  Both molecules may be represented by the symbol 

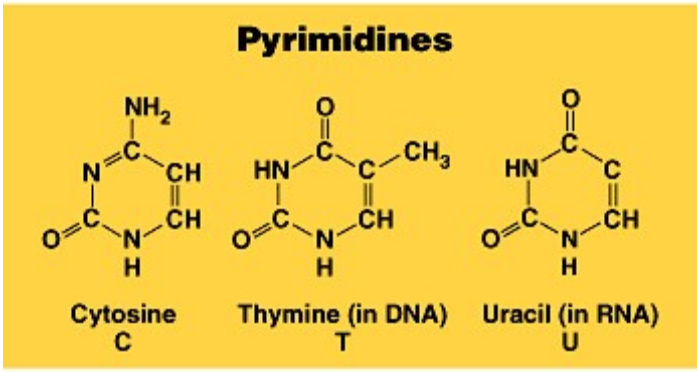
Nitrogenous bases
There are 2 types:
Purines:

- Having Two ring structure
- Adenine (A) and Guanine (G)

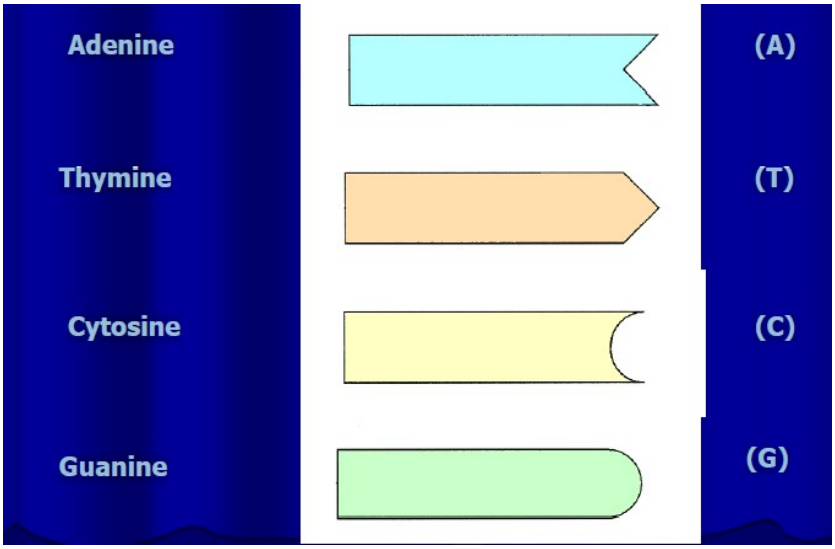


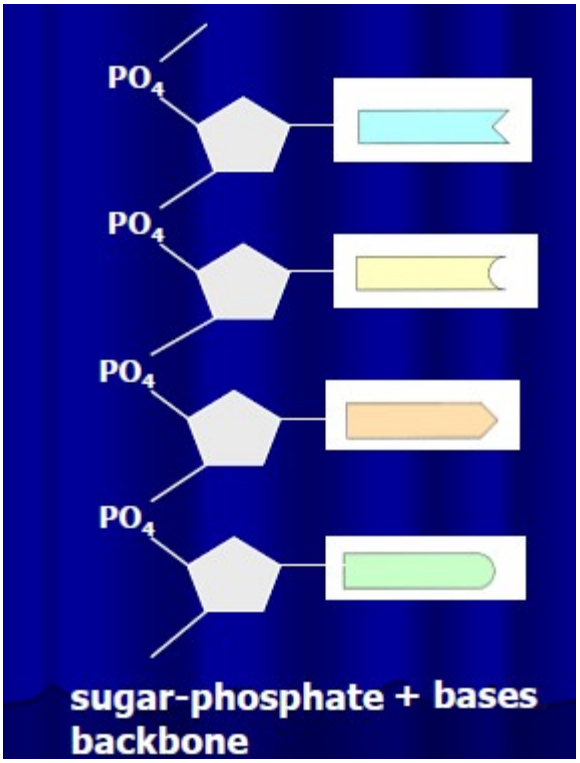
Pyrimidines:

- Single ring structure
- Cytosine (C) and Thymine (T) or Uracil (U).



The organic bases were represented by the following symbols





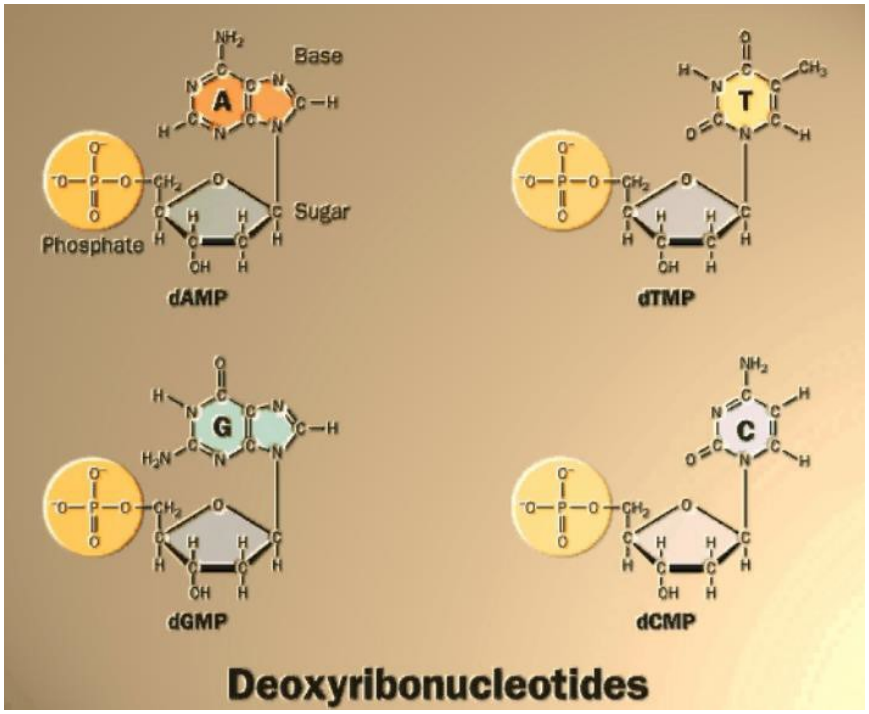
A molecule of DNA is formed by millions of nucleotides joined together in a long chain

DNA

DNA stands for deoxyribose nucleic acid This chemical substance is present in the nucleus of all cells in all living organisms. DNA controls all the chemical changes which take place in cells .The kind of cell which is formed, (muscle, blood, nerve etc) is controlled by DNA

The kind of organism which is produced (buttercup, giraffe, herring, human etc) is controlled by DNA.

- Pentose Sugar is deoxyribose (no OH at 2' position)
- Bases are Purines (A, G) and Pyrimidine (C, T).

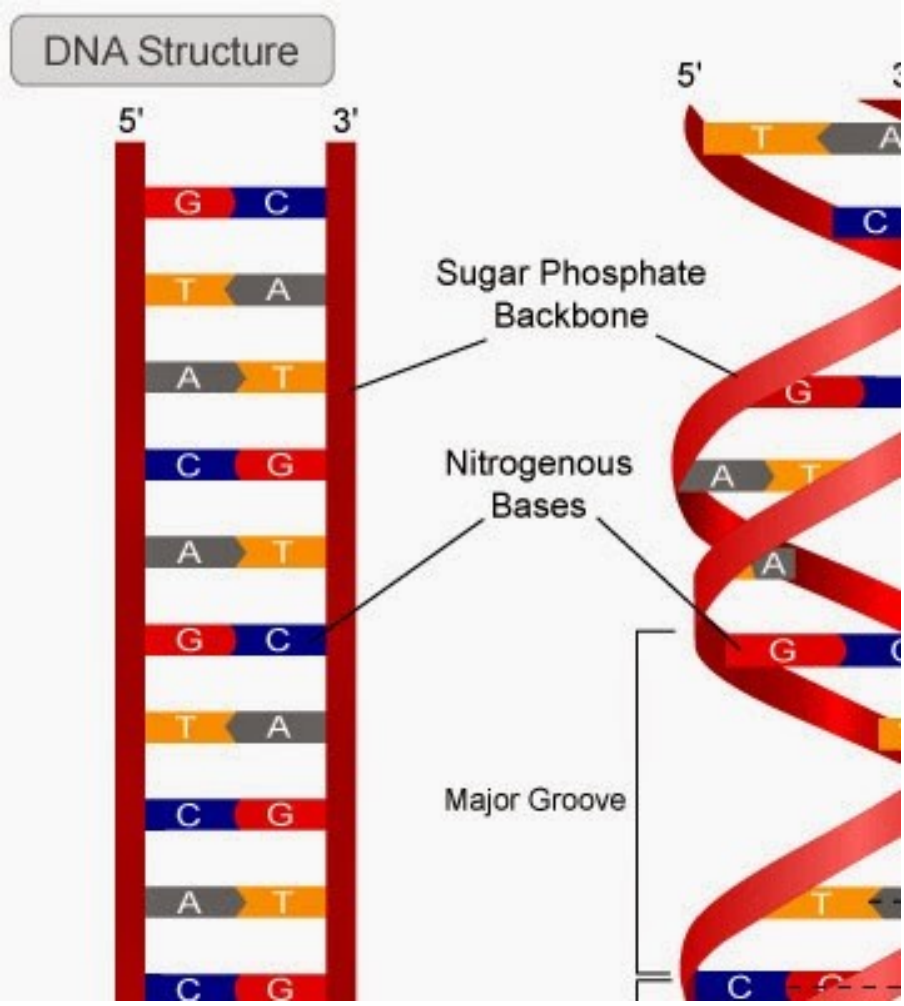


- **DNA structure was proposed by James Watson** (1928 -) and **Francis Crick**
- **Deoxyribonucleic Acid** (**DNA**) is a **double-stranded double helical structured molecule**.
- It consists of two **sugar-phosphate backbones** on the outside, held together by **hydrogen bonds** between pairs of **nitrogenous bases** on the inside. The bases are of four types (**A**, **C**, **G**, & **T**): pairing always occurs between **A** & **T**, and **C** & **G**
- D.N.A is a macromolecule made up of helically twisted two anti-parallel strands.

  (One strand moves from 3'-5' direction and other moves from 5'-3' direction).

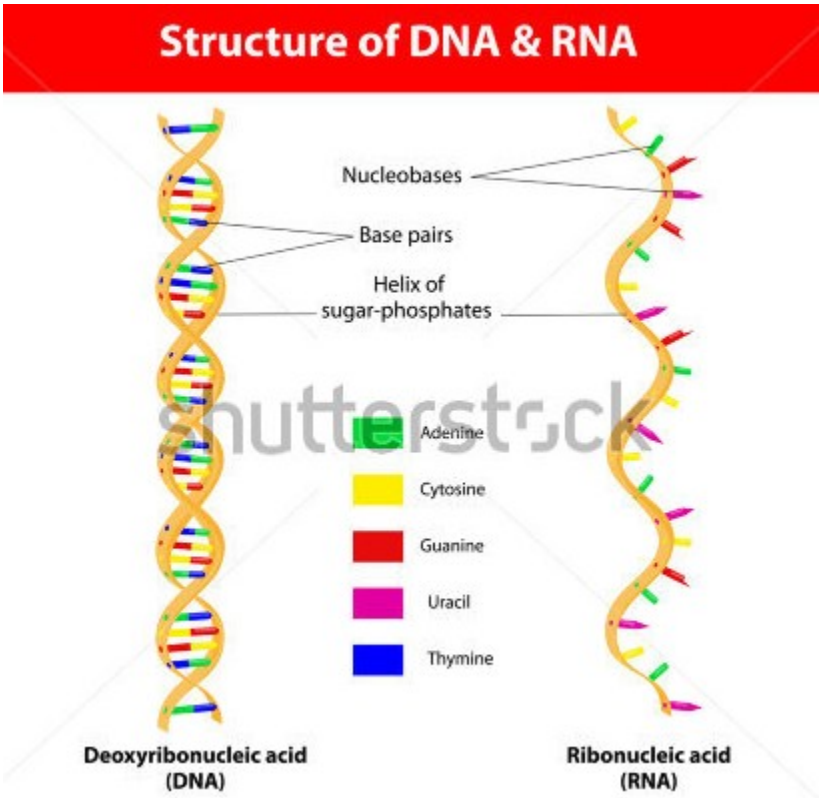
- The D.N.A has got many complete turns. The length of one complete turn is3.4nm and has about 10 nitrogen base pairs. The distance between two base pairs is 0.34nm.

- A purine base is attached to its specific pyrimidine base. I.e:Adenine combines to Thymine by double hydrogen bond. (A=T) and Cytosine combines to Guanine by triple hydrogen bond. (G≡C)



RNA

- RNA is single stranded
- The pentose sugar is Ribose so called Ribonucleic acid(RNA)
- It is also known as polyribonucleotides as it is formed by the polymer of few to several hundred ribonucleotides by a phosphodiester bond.
- Each ribonucleotide has got phosphate group, ribose as pentose sugar and nitrogen base like in a D.N.A.
- The nitrogen bases such as Adenine(A) and Guanine(G) of purine and Cytosine(C) and Uracil(U) of pyrimidine are present. The Thymine of pyrimidine is altogether absent.
- The strand or backbone of R.N.A is formed by ribose sugar and phosphate group by phosphodiester bonds.



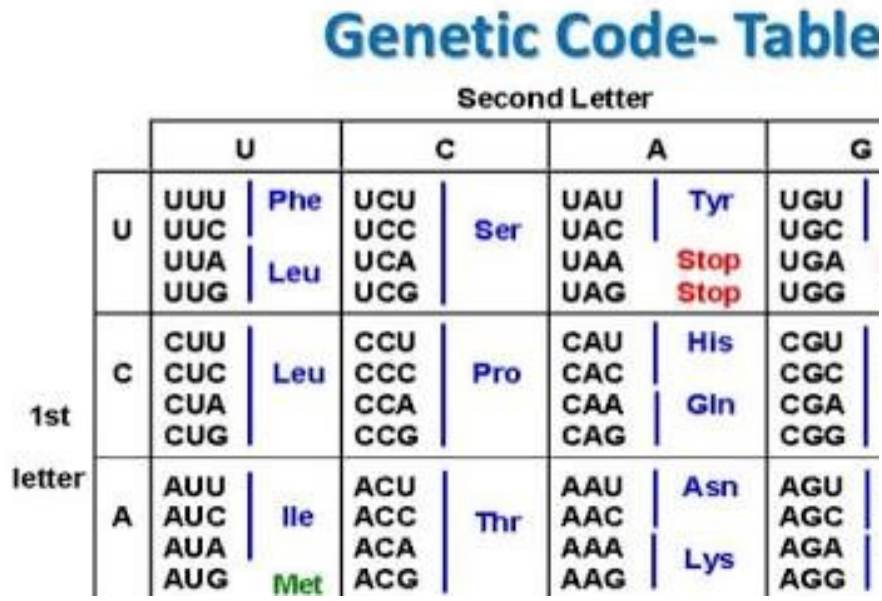
Genetic code

The genetic code is the sequence of nucleotide bases in nucleic acids (DNAand RNA) that code for amino acidchains in proteins. DNA consists of the four nucleotide bases: adenine (A), guanine (G), cytosine (C) and thymine (T). RNA contains the nucleotides adenine, guanine, cytosine and uracil (U). When three continuous nucleotide bases code for an amino acid or signal the beginning or end of protein synthesis, the set is known as a codon. These triplet sets provide the instructions for the production of amino acids. Amino acids are linked together to form proteins.

Codons

RNA codons designate specific amino acids. The order of the bases in the codon sequence determines the amino acid that is to be produced. Any of the four nucleotides in RNA may occupy one of three possible codon positions. Therefore, there are 64 possible codon combinations. Sixty-one codons specify amino acids and three **(UAA, UAG, UGA)** serve as **stop signals** to designate the end of protein synthesis. The codon **AUG** codes for the amino acid **methionine** and serves as a **start signal** for the beginning of translation. Multiple codons may also specify the same amino acid. For example, the codons UCU, UCC, UCA, UCG, AGU, and AGC all specify serine. The RNA codon table above lists codon combinations and their designated amino acids. Reading the table, if uracil (U) is in the first codon position, adenine (A) in the second,

and cytosine (C) in the third, the codon UAC specifies the amino acid tyrosine. The abbreviations and names of all 20 amino acids are listed below.



Amino Acids

Ala: Alanine **Asp:** Aspartic acid **Glu:** Glutamic acid **Cys:** Cysteine
Phe: Phenylalanine **Gly:** Glycine **His:** Histidine **Ile:** Isoleucine
Lys: Lysine **Leu:** Leucine **Met:** Methionine **Asn:** Asparagine
Pro: Proline **Gln:** Glutamine **Arg:** Arginine **Ser:** Serine
Thr: Threonine **Val:** Valine **Trp:** Tryptophan **Tyr:** Tyrosine

Protein Production

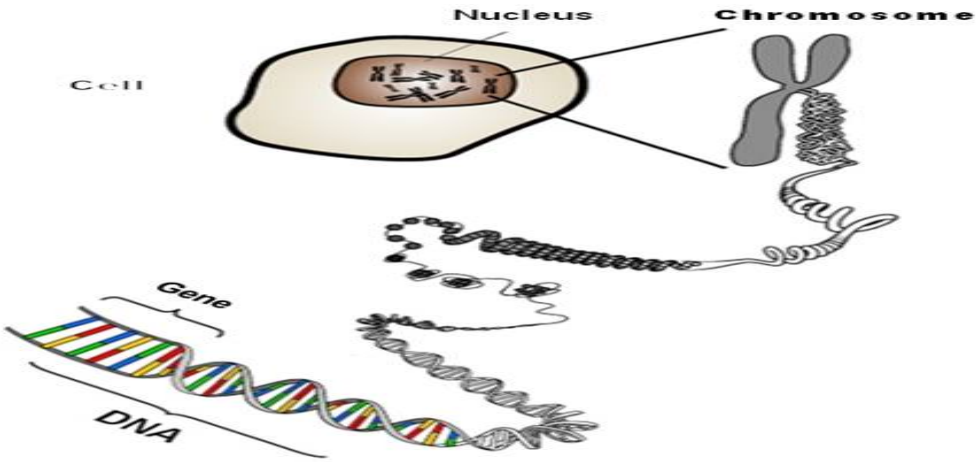
Proteins are produced through the processes of DNA transcription and translation. The information in DNA is not directly converted into proteins, but must first be copied into RNA. DNA transcription is the process in protein synthesis that involves the transcribing of genetic information from DNA to RNA. Certain proteins called transcription factors unwind the DNA strand and allow the enzyme RNA polymerase to transcribe only a single strand of DNA into a single stranded RNA polymer called messenger RNA (mRNA). When RNA polymerase transcribes the DNA, guanine pairs with cytosine and adenine pairs with uracil.

Since transcription occurs in the nucleus of a cell, the mRNA molecule must cross the nuclear membrane to reach the cytoplasm. Once in the cytoplasm, mRNA along

with ribosomes and another RNA molecule called transfer RNA, work together to translate the transcribed message into chains of amino acids. During translation, each RNA codon is read and the appropriate amino acid is added to the growing polypeptide chain. The mRNA molecule will continue to be translated until a termination or stop codon is reached.

Genes and Evolution

- A gene is a heritable factor that consists of a length of DNA and influences a specific characteristic.
- A gene occupies a specific position on a chromosome.



Chromosomes are made from DNA. Genes are short sections of DNA. Alleles are different forms of a gene. They can be dominant or recessive. Genetic diagrams help us to understand the possible outcomes when parents produce offspring. Cystic fibrosis is a disorder of the cell membranes caused by a recessive allele. Sickle cell disease is another example of a genetic disorder.

DNA, genes and chromosomes

DNA

DNA (deoxyribose nucleic acid) molecules are large and complex. They carry the genetic code that determines the characteristics of a living thing.

Except for identical twins, each person's DNA is unique. This is why people can be identified using DNA fingerprinting. DNA can be cut up and separated, forming a sort of 'bar code' that is different from one person to the next.

Genes

A *gene* is a short section of DNA. Each gene codes for a specific protein by specifying the order in which amino acids must be joined together.

Chromosomes

The cell's nucleus contains *chromosomes* made from long DNA molecules.

The above diagram shows the relationship between the cell, its nucleus, chromosomes in the nucleus, and genes.

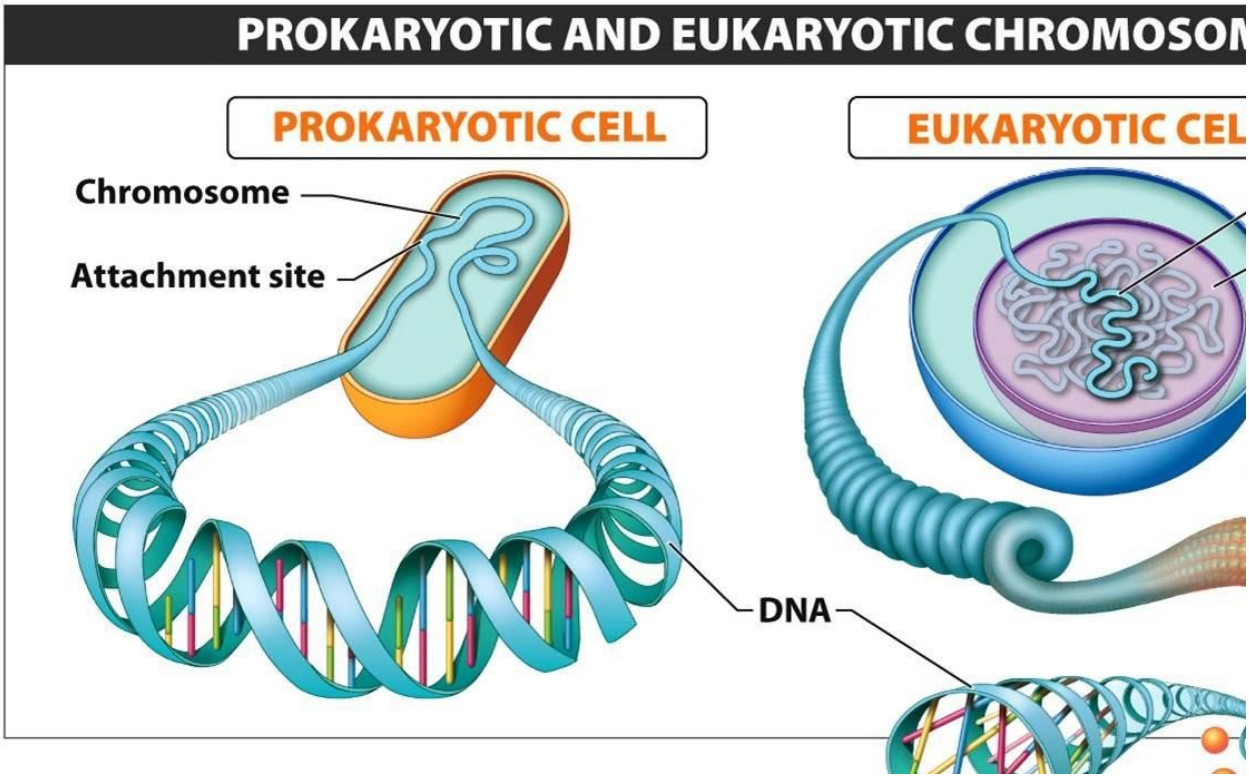
Alleles

Some characteristics, such as eye colour and the shape of the earlobe, are controlled by a single gene. These genes may have different forms.

Different forms of the same gene are called *allele*s (pronounced al-eels). The gene for eye colour has an allele for blue eye colour and an allele for brown eye colour.

Alleles are *dominant* or *recessive*:

- the characteristic controlled by a **dominant** allele develops if the allele is present on **one or both** chromosomes in a pair
- the characteristic controlled by a **recessive** allele develops only if the allele is present on **both** chromosomes in a pair



Genetic Expression

Each person's chromosomes contain a unique catalog of genes, the **genotype** for that person. The expression of those genes yields observable traits known as

the **phenotype.** Though the phenotype is what is seen, the genotype is responsible for the inheritance and expression of those traits.

GENOME EVOLUTION

Genome evolution is the process by which a genome changes in structure (sequence) or size over time.

• mechanisms for evolution

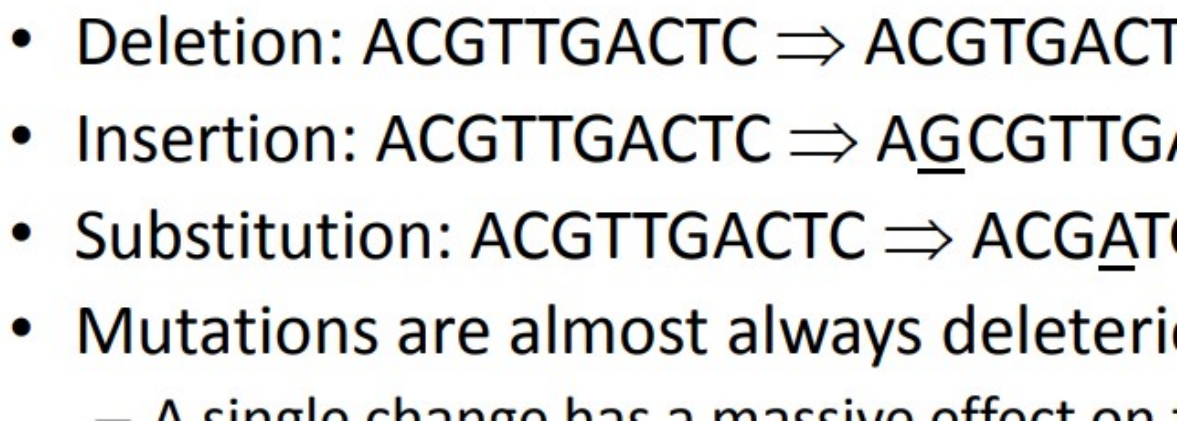
1. inheritance
2. mutation
3. Pseudogenes
4. Gene duplication
5. Genome reduction and gene loss
6. Transposable elements

1. Inheritance:

- properties from parents
- Embryo has cells with 23 pairs of chromosomes
- Each pair: 1 chromosome from father, 1 from mother

2.Mutation:

- Genes alter (slightly) during reproduction
- Caused by errors, from radiation, from toxicity
- 3 possibilities: deletion, insertion, **alteration**



3.Pseudogenes

• There are many mechanisms by which a functional gene can become a pseudogene including the deletion or insertion of one or multiple nucleotides.

• This can result in a shift of reading frame, causing the gene to no longer code for the expected protein, introduce a premature stop codon or a mutation in the promoter region.

4.Exon shuffling

• Exon shuffling is a mechanism by which new genes are created.

• This can occur when two or more exons from different genes are combined together or when exons are duplicated.

• Exon shuffling results in new genes by altering the current intron-exon structure.

5.Genome reduction and gene loss

• Many species exhibit genome reduction when subsets of their genes are not needed anymore.

• e.g. when their nutrients are supplied by a host. As a consequence, they lose the genes needed to produce these nutrients.

6.Gene duplication

• process by which a region of DNA coding for a gene is duplicated.

• This can occur as the result of an error in recombination

• Duplicate genes are often immune to the selective pressure under which genes normally exist.

7.Transposable elements

• Transposable elements are regions of DNA that can be inserted into the genetic code through one of two mechanisms.

• These mechanisms work similarly to "cut-andpaste" and "copy-and-paste" functionalities in word processing programs.