Course 01: Introduction to Bioinformatics

1. Definition of bioinformatics

According to the National Center of Biotechnology Informatics Informatics NCBI: science in which biology , computer sciences and information technology merge into a single discipline.

" there Bioinformatics is the science in which biology, computer science, and technology merge into a single science.

Bioinformatics , newly included in biological teaching systems (it emerged in the 1980s). It is a recent discipline and a multidisciplinary research field where biologists, physicians, computer scientists , mathematicians and physicists work together to solve a scientific problem posed by biology.

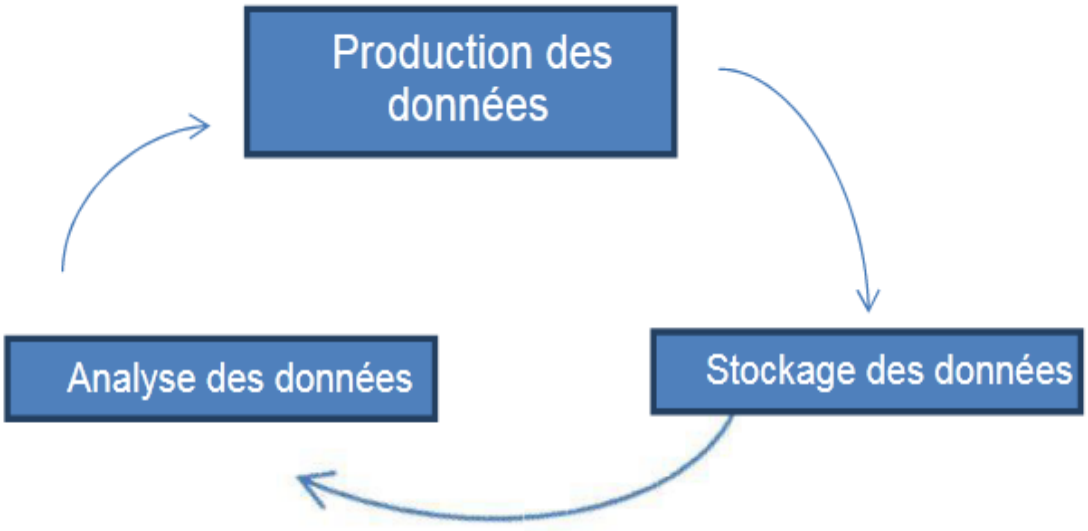
It can also be defined as the discipline of "in silico" analysis of the biological information contained in nucleic and protein sequences.

Three main activities:

• Acquisition and organization of biological data

• Software design for data analysis, comparison, and modeling

• Analysis of the results produced by the software.



2. **Goals :**

It has become the ultimate tool for:

•interpret biomolecular data,

•analyze the structure of molecules,

• Compare this structure to the other existing molecules in biological databases,

•predict the role and function of this structure, …

3. **Scope and use:**

She is interested in the data from:

- genome (the entire genetic material of the cell),

- transcriptome (transcribed mRNAs), - transcriptome (transcribed mRNAs),

- proteome (the set of biosynthesized proteins),

- metabolome (organic molecules such as lipids, carbohydrates, which are part of the metabolic activities of the living cell).

A reminder in molecular biology

1. Deoxyribonucleic acid (DNA)

- DNA is the carrier of genetic information.

- DNA is a long molecule, made of two strands that twist into a double helix.

- The two strands of the double helix suggest a DNA replication mechanism

- Each strand carries a sequence of nucleotides

- Four types of nucleotides: (Adenine, Cytosine, Guanine, Thymine).

- The genomic text is written in a 4-letter alphabet: A, C, G, T.

These two strings have three properties:

• Antiparallels:

This means that the two strands of nucleotides are parallel but in opposite directions.

For one strand, the 5' to 3' direction happens to be, for example, from top to bottom and for the 2nd strand, the 5' to 3' direction will be the reverse, from bottom to top.

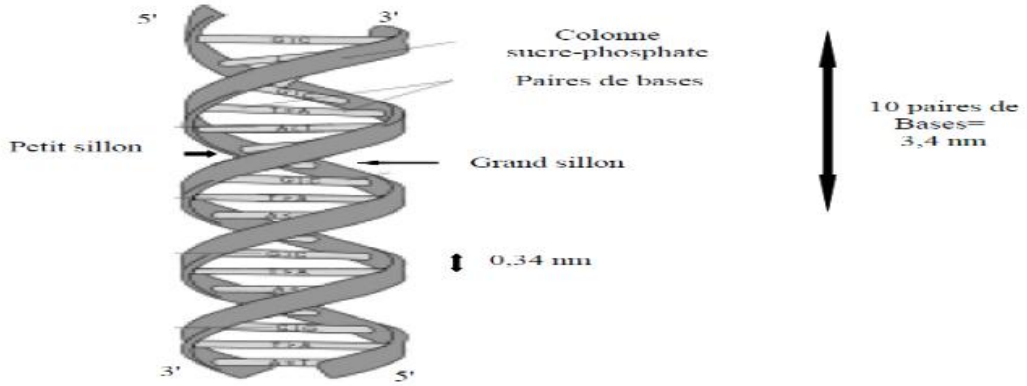
• Complementary

The complementarity rule is as follows: opposite A we have T and opposite C we have G. Indeed, the distribution of nitrogenous bases in DNA is not arbitrary, the number of hydrogen bonds for the A - T pair is two and for the C - G pair is three.

• Helical

This equal distribution of A and T on the one hand, and of G and C on the other, therefore implies a structure

particularity of DNA, structure which has greatly intrigued biologists.



Représentation de la double hélice d'ADN

The structure of DNA represented by the figure, the two strands are oriented, we speak of the 5' → 3' orientation by biochemical considerations.

- The 5' end terminates with a phosphate group.

- The 3' end terminates with a hydroxyl group.

It should be remembered that there is a strand oriented in the direction 5' → 3' and a strand oriented against the direction 3' → 5'.

Figure: Schematic representation of a sequence of nucleotides in a fragment of double-stranded DNA .

DNA **is capable of replicating, meaning it can copy itself faithfully** . DNA splits into two strands, each separated strand serving as a template for building a complementary strand. The result: two new DNA molecules, each with one old and one new strand.

DNA serves three main functions: **the formation of proteins and RNA, the exchange of genetic material during cell division during meiosis, and the facilitation of genetic mutations within a population** .

Nucleotides **are** the basic building blocks of nucleic acids, DNA and RNA. They **are** composed of three main elements: a nitrogenous base, linked to a sugar molecule and a phosphate group.

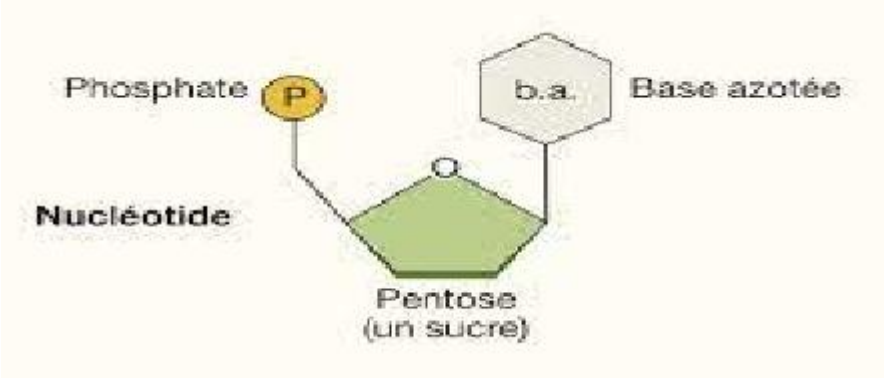


Figure: The components of a nucleotide

2. Ribonucleic acid (RNA):

RNAs are polymers of nucleotides; their general structure is very similar to that of DNA, but three essential differences should be noted:

RNA is single-stranded, the sugar is ribose, and there is no thymine in the RNA structure; it is replaced by uracil.

RNA molecules are made up of a single polynucleotide chain . Except in some viruses, where pairings by double helices can be established between A and U on the one hand and between G and C on the other, but these pairings are formed between 2 bases of the same chain and have no regularity character unlike DNA.

The cell essentially contains 4 types of RNA:

rRNA (ribosomal), tRNA (transfer), mRNA (messenger) and snRNA (smallnuclear).

rRNAs are the most abundant (82%), followed by tRNAs (16%) and mRNAs

(~2%).

2.1. Ribosomal RNAs (rRNA):

Ribosomes are the particles necessary for protein synthesis. Located in the cytosol, ribosomes are veritable protein factories within the cell. Ribosomes are also found in mitochondria. rRNAs are the major constituents of ribosomes, making up approximately 65%.

rRNAs are components of the ribosome (necessary for protein synthesis). A functional ribosome is itself composed of two subunits, each made up of a mixture of proteins (r-proteins) and RNA (rRNA). Ribosomes are located in the cytoplasm and are essential for protein synthesis. They are veritable " protein factories."

2.2. Transfer RNAs (tRNAs):

They are called tRNA because they transfer and transport amino acids from the cytoplasm to the ribosome, the site of protein synthesis. A tRNA molecule has the general structure of RNA. The tRNA chain folds to give it a cloverleaf shape. Two sites are important in a tRNA molecule :

- The 3'OH end will be where the amino acid to be transported will be attached

- The anticodon (triplet) located on a loop of the tRNA will play a very important role because it will recognize the codon of the mRNA. This anticodon-codon pairing occurs in an antiparallel and complementary manner between the bases of the codon and the anticodon.

2.3. Messenger RNA

It is composed of a single chain of nucleotides containing the same types of bases, AUCG. It is called messenger RNA (mRNA) because it carries the genetic information contained in the DNA to the ribosome, where protein synthesis takes place. The size of the mRNA molecule depends on the length of the polypeptide chain(s) it codes for. mRNAs are renewed very quickly; they are rapidly produced and rapidly degraded. They only last for the duration of a single message. However, an mRNA can be read multiple times by the ribosome.

2.4. Small nuclear RNAs (snRNAs)

• SnRNA (small nuclear) the smallest RNA molecules, we will see later that these RNAs play an important role in the maturation of Pre-mRNAs.

3. The genome

The genome, that is, the complete set of genes of an organism, is structurally defined by the chromosomes within each cell, according to the chromosomal theory of heredity.

Each chromosome consists of a DNA molecule (Figure 5), the physical support of the genes, and associated proteins.

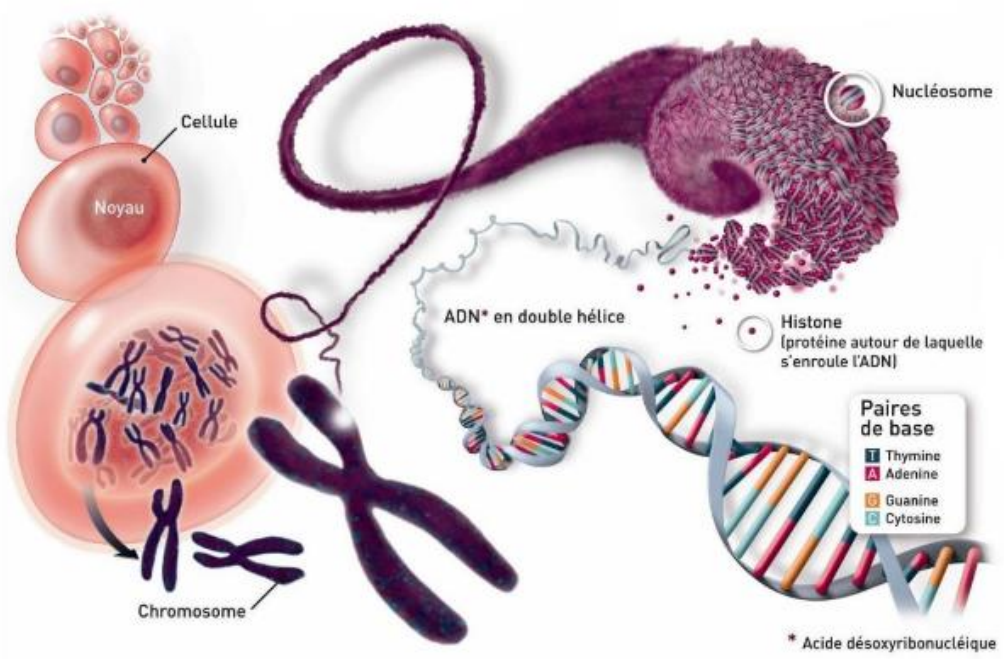


Figure: Diagram of the human genome, showing the double helix structure of DNA. (Source: Premamag.wordpress.com).

3.1. The prokaryotic genome

The genetic material of bacteria is not organized in the same way as in eukaryotes. However, bacterial genes are arranged linearly on the chromosome. Prokaryotes generally possess a single circular chromosome, often in a single copy. Nevertheless, with the use of pulsed-field gel electrophoresis, which allows the separation of large DNA molecules, it has been shown that some bacteria have a linear chromosome or several circular and linear chromosomes.
In addition, prokaryotes possess DNA in an extra-genomic form. This is a small circular DNA molecule, called a plasmid, capable of replicating independently of the chromosome.

3.2. The eukaryotic genome

In eukaryotes, genomes are visualized as filamentous, non-circular structures located primarily in the nucleus, which can exhibit variable configurations depending on the cell cycle (see mitosis). There are also mitochondrial and chloroplast chromosomes, which are mostly circular. These are smaller than nuclear chromosomes and do not have a filamentous appearance. The genes located on these

extranuclear chromosomes do not follow the laws of Mendelian inheritance .

4. Genomics is a branch of modern biology . It studies the functioning of an organism, an organ, a cancer , etc., at the genome level , rather than being limited to the level of a single gene .

Genomics is divided into two branches:

- **Structural genomics :** which is responsible for sequencing the entire genome;
- **Functional genomics :** which aims to determine the function and expression of sequenced genes by characterizing the transcriptome and proteome .

- The **transcriptome is the complete set of RNA molecules produced by the transcription of the genome.** Transcriptomic analysis can characterize the transcriptome of a particular tissue or cell type, or compare transcriptomes under different experimental conditions.
- The **proteome** is the set of proteins expressed in a cell , part of a cell (membranes , organelles) or group of cells (organ, organism, group of organisms) under given conditions and at a given time.