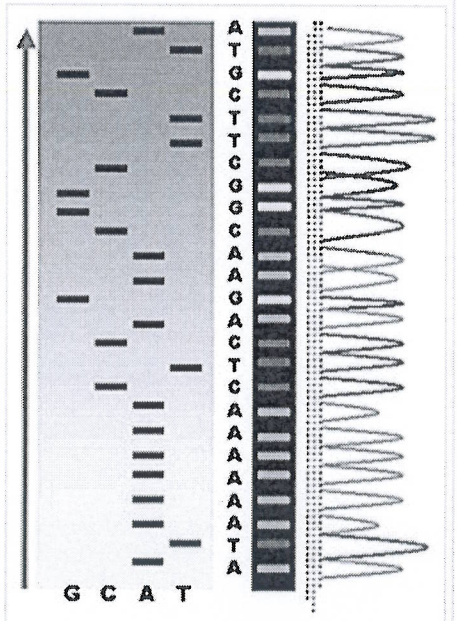
DNA sequencing

From Wikipedia, the free encyclopedia

DNA sequencing is the process of determining the precise order of nucleotides within a DNA molecule. It includes any method or technology that is used to determine the order of the four bases—adenine, guanine, cytosine, and thymine—in a strand of DNA. The advent of rapid DNA sequencing methods has greatly accelerated biological and medical research and discovery.

Knowledge of DNA sequences has become indispensable for basic biological research, and in numerous applied fields such as medical diagnosis, biotechnology, forensic biology, virology and biological systematics. The rapid speed of sequencing attained with modern DNA sequencing technology has been instrumental in the sequencing of complete DNA sequences, or genomes of numerous types and species of life, including the human genome and other complete DNA sequences of many animal, plant, and microbial species.

The first DNA sequences were obtained in the early 1970s by academic researchers using laborious methods based on two-dimensional chromatography. Following the development of fluorescence-based sequencing methods with automated analysis,[1] DNA sequencing has become easier and orders of magnitude faster.[2]



An example of the results of automated chain-termination DNA sequencing.

Contents

- Molecular biology – studying the genome itself, how proteins are made, what proteins are made, identifying new genes and associations with diseases and phenotypes, and identifying potential drug targets
- Evolutionary biology – studying how different organisms are related and how they evolved
- Metagenomics – Identifying species present in a body of water, sewage, dirt, debris filtered from the air, or swab samples of organisms. Helpful in ecology, epidemiology, microbiome research, and other fields.

Less-precise information is produced by non-sequencing techniques like DNA fingerprinting. This information may be easier to obtain and is useful for:

- Detecting the presence of known genes for medical purposes (see genetic testing)
- Forensic identification
- Parental testing

The four canonical bases

The canonical structure of DNA has four bases: Thymine (T), Adenine (A), Cytosine (C), and Guanine (G). DNA sequencing is the determination of the physical order of these bases in a molecule of DNA. However, there are many other bases that may be present in a molecule. In some viruses (specifically, bacteriophage), cytosine may be replaced by hydroxy methyl or hydroxy methyl glucose cytosine.[3] In mammalian DNA, variant bases with methyl groups or phosphosulfate may be found.[4][5] Depending on the sequencing technique, a particular modification may or may not be detected, e.g., the 5mC (5 methyl cytosine) common in humans may or may not be detected.[6]

History

RNA sequencing

Though the structure of DNA was established as a double helix in 1953,[7] several decades would pass before fragments of DNA could be reliably analyzed for their sequence in the laboratory. RNA sequencing was one of the earliest forms of nucleotide sequencing. The major landmark of RNA sequencing is the sequence of the first complete gene and the complete genome of Bacteriophage MS2, identified and published by Walter Fiers and his coworkers at the University of Ghent (Ghent, Belgium), in 1972[8] and 1976.[9]

Early DNA sequencing methods

The first method for determining DNA sequences involved a location-specific primer extension strategy established by Ray Wu at Cornell University in 1970.[10] DNA polymerase catalysis and specific nucleotide labeling, both of which figure prominently in current sequencing schemes, were used to