Popular Article

TRENDS IN AGRICULTURE SCIENCE

Vol.3 Issue 10 October 2024, Page: 2261-2267

DNA Sequencing

R. Muthuvijayaragavan[1*] and K. Krishna Surendar[2]

[1]Department of Plant Biotechnology, Centre for Molecular Biology and Biotechnology, Tamil Nadu Agricultural University, Coimbatore, Tamil Nadu, India – 641 003

[2]Department of Crop Physiology, Directorate of Crop Management, Tamil Nadu Agricultural University, Coimbatore, Tamil Nadu, India – 641 003

DOI:10.5281/TrendsinAgri.14019267

Introduction

DNA double helix structure was first explained by Watson and Crick. The first natural polynucleotide DNA sequence was announced in 1953. Due to the development of many sequencing techniques, the entire human genome was sequenced in 2003 as a result of the Human Genome Project. The goal of the Genome Project, a large-scale scientific initiative, is to investigate and evaluate the chemical sequence of the 50 000-100 000 genes that make up the human genome or the entire collection of all genetic material. The separation and examination of the genetic code included in DNA provide the basis for this huge project. Considering the size of the human genome, scientists must have developed new techniques for DNA analysis that can quickly, cheaply, and reliably analyse massive amounts of data. DNA sequencing methods requiring large- scale application have driven technology to both enhance capacity and decrease instrument size. This demand has prompted the creation of automated equipment that speeds up and lowers the cost of biochemical processes connected to sequencing, enhances the analysis of these reactions, and makes it simpler to enter the resulting data into databases (Behjati and Tarpey 2013). It was not possible for scientists at the time to sequence nearly a full gene because they could only sequence a few base pairs per year. Despite difficulties in the sequencing process, the first complete genome sequencing was completed with an explosion of RNA and DNA sequencing that improved procedures and provided new data. This eventually gave rise to the Maxam and Gilbert chemical degradation DNA sequencing method, which physically separates terminally tagged DNA fragments by electrophoresis and chemically cleaves specific bases of those pieces (Collings *et al.,* 1998). Maxam-Gilbert, called the first-generation sequencing method, is based on chemical fragmentation of DNA and imaging with electrophoresis (Dogan *et al.,* 2017). Later, a new technique with higher simplicity, reliability, and a lower hazardous level was developed. It was simply called the Sanger sequencing method. Automated Sanger DNA sequencing with fluorescent dye labels predominated when the Human

Published: 30 October 2024
ISSN: 2583-7850

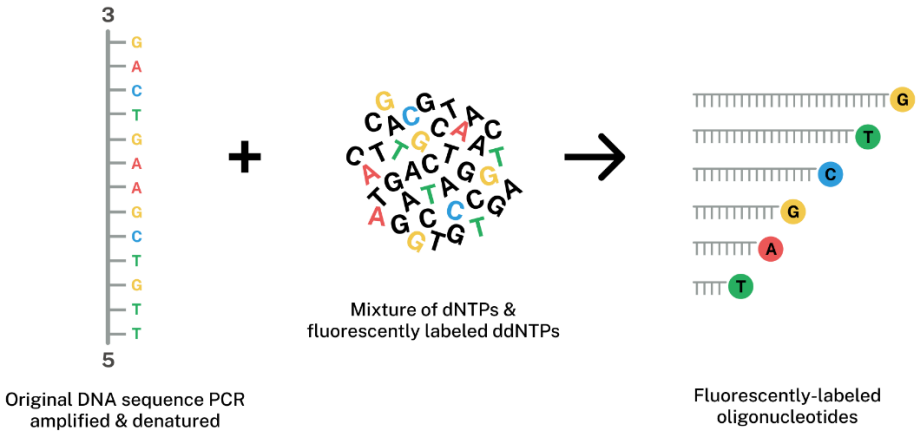
Genome Project was finished in 2003 (Behjati and Tarpey 2013). Sanger technique, which uses fewer chemicals and is less harmful, is still regarded as the gold standard today (Zhang *et al.,* 2021).

DNA sequencing refers to methods for determining the order of the nucleotide bases adenine, guanine, cytosine, and thymine in a molecule of DNA. The first DNA sequence was obtained by academic researchers, using laboratories methods based on 2- dimensional chromatography in the early 1970s. By the development of dye based sequencing method with automated analysis, DNA sequencing has become easier and faster.

Sequencing Methods

1. **The Chemical Method** (also called the Maxam–Gilbert method after its inventors).
2. **The Chain Termination Method** (also known as the Sanger dideoxy method after its inventor).

The Maxam–Gilbert technique depends on the relative chemical liability of different nucleotide bonds, whereas the Sanger method interrupts elongation of DNA sequences by incorporating dideoxynucleotides into the sequences. The chain termination method is the method more usually used because of its speed and simplicity.



Chemical Cleavage Method (Maxam–Gilbert Method)

In 1976-1977, Allan Maxam and Walter Gilbert developed a DNA sequencing method based on chemical modification of DNA and subsequent cleavage at specific bases. It requires radioactive labelling at one end and purification of the DNA fragment to be sequenced. Chemical treatment generates breaks at a small proportion of one or two of the four nucleotide based in each of four reactions (G, A+G, C, C+T). A series of labelled fragments is generated, from the radiolabelled end to the first 'cut' site in each molecule. The fragments in the four reactions are arranged side by side in gel electrophoresis for size separation. To visualize the fragments, the gel is exposed to X-ray film for autoradiography, yielding a series of dark bands each corresponding to a radiolabelled DNA fragment, from which the sequence may be inferred.

Key Features	Advantages	Disadvantages

Published: 30 October 2024
ISSN: 2583-7850

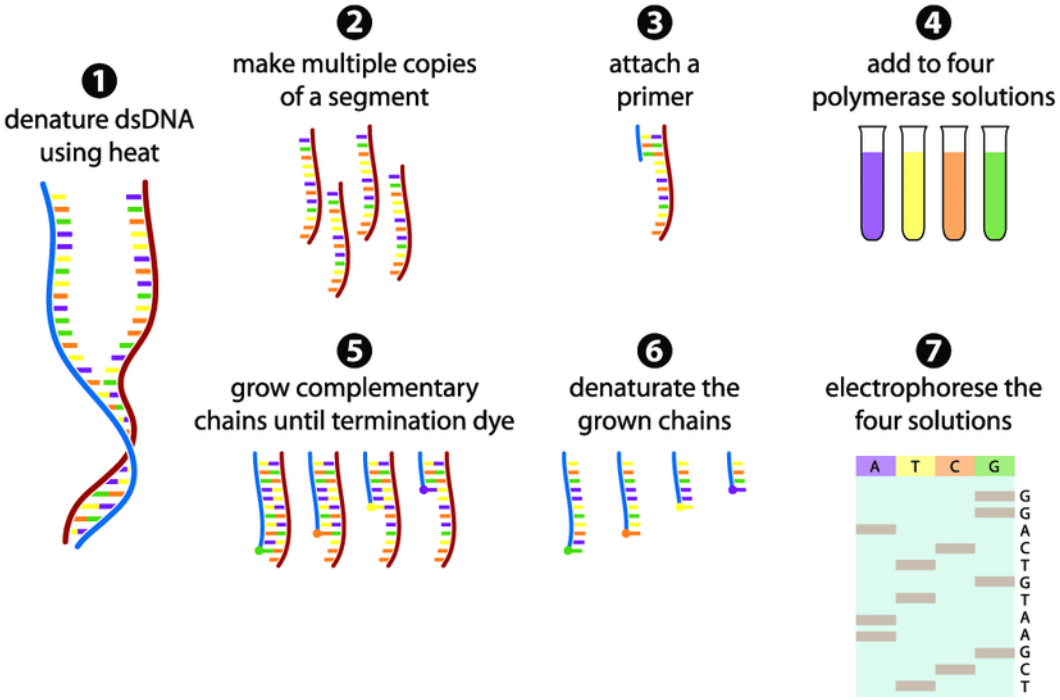
• Base-specific cleavage of DNA by certain chemicals • Four different chemicals, one for each base • A set of DNA fragments of different sizes • DNA fragments contain up to 500 nucleotides	• Purified DNA can be read directly • Homopolymeric DNA runs are sequenced as efficiently as heterogeneous DNA sequences • It can be used to analyse DNA protein interactions (*i.e.* foot printing) • It can be used to analyse nucleic acid structure and epigenetic modifications to DNA	• It requires extensive use of hazardous chemicals. • It has a relatively complex set up / technical complexity. • It is difficult to "scale up" and cannot be used to analyse more than 500 base pairs. • The read length decreases from incomplete cleavage reactions. • It is difficult to make Maxam-Gilbert sequencing, based DNA kits.

Chain Termination Method (Sanger Dideoxy Method)

The chain terminator method is more efficient and uses fewer toxic chemicals and lower amount of radioactivity than the method of Maxam and Gilbert. The key principle of the Sanger method was the use of dideoxynucleotide triphosphates (ddNTPs) as DNA chain terminators. It requires a single-stranded DNA template, a DNA primer, a DNA polymerase, radioactively or fluorescently labelled nucleotides and modified nucleotides that terminate DNA strand elongation. The DNA sample is divided into four separate sequencing reactions, containing all four of the standard deoxynucleotides (dATP, dGTP, dCTP, dTTP) and the DNA polymerase. To each reaction is added only one of the four dideoxynucleotide (ddATP, ddGTP, ddCTP, ddTTP) which are the chain terminating nucleotides, lacking a 3'-OH group required for the formation of a phosphodiester bond between two nucleotides, thus terminating DNA strand extension and resulting in DNA fragments of

Published: 30 October 2024
ISSN: 2583-7850

varying length. The newly synthesized and labelled DNA fragments are heat denatured, and separated by size by gel electrophoresis on a denaturing polyacrylamide-urea gel with each of the four reactions run in one of the four individual lanes (lanes A, T, G, C). The DNA bands are then visualized by autoradiography or UV light, and the DNA sequence can be directly read off the X-ray film or gel image. A dark band in a lane indicates a DNA fragment that is result of chain termination after incorporation of a dideoxynucleotide (ddATP, ddGTP, ddCTP, or ddTTP). The relative position of the different bands among the four lanes are then used to read (from bottom to top) the DNA sequence. The technical variations of chain termination sequencing include tagging with nucleotides containing radioactive phosphorus for labelling, or using a primer labelled at the 5' end with a fluorescent dye. Dye- primer sequencing facilitates reading in an optical system for faster and more economical analysis and automation.



Key Features	Advantage	Limitations
• Uses dideoxy nucleotides to terminate DNA synthesis. • DNA synthesis reactions in four separate tubes • Radioactive dATP is also included in all the tubes so the DNA products will be radioactive.	Chain termination methods have greatly simplified DNA sequencing.	• Non-specific binding of the primer to the DNA, affecting accurate read-out of the DNA sequence. • DNA secondary structures affecting the fidelity of the sequence.

Published: 30 October 2024
ISSN: 2583-7850

• Yielding a series of DNA fragments whose sizes can be measured by electrophoresis.		

Significance of DNA Sequencing

Information obtained by DNA sequencing makes it possible to understand or alter the function of genes. DNA sequence analysis demonstrates regulatory regions that control gene expression and genetic "hot spots" particularly susceptible to mutation. Comparison of DNA sequences facilitate identification of conserved regions, which are useful for development of specific hybridization probes to detect microorganisms including viruses in clinical samples. DNA sequencing has become sufficiently fast and inexpensive to allow laboratory determination of microbial sequences for identification of microbes. Sequencing of the 16S ribosomal subunit can be used to identify specific bacteria. Sequencing of viruses can be used to identify the virus and distinguish different strains. DNA sequencing shows gene structure that helps research workers to find out the structure of gene products.

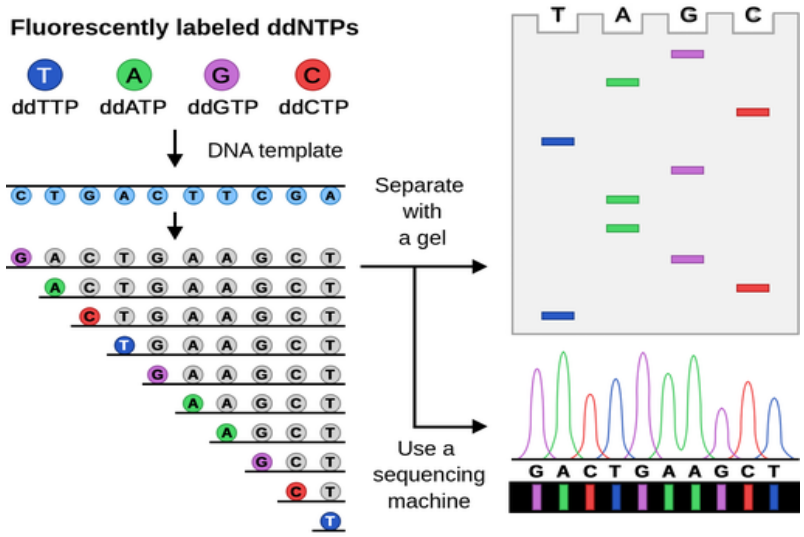


Table 1. Landmarks in DNA Sequencing

S. No.	Year	Salient features
1.	1953	Discovery of the structure of the DNA double helix
2.	1972	Development of recombinant DNA technology
3.	1977	The first complete genome of bacterio- phage uX174 sequenced
4.	1977	Allan Maxam and Walter Gilbert publish "DNA sequencing by chemical degradation"

Published: 30 October 2024
ISSN: 2583-7850

5.	1984	Medical Research Council scientists decipher the complete DNA sequence of the Epstein-Barr virus, 170 kb
6.	1986	Leroy E. Hood's laboratory at the California Institute of Technology and Smith announced the first semi-automated DNA sequencing machine
7.	1987	Applied Biosystems marketed first automated sequencing machine, the model ABI 370
8.	1990	The U.S. National Institutes of Health (NIH) begins large-scale sequencing trials on *Mycoplasma capricolum*, *E. coli, C. elegans*, and *S. cerevisiae*
9.	1991	Sequencing of human expressed sequence tags begins in Craig Venter's lab
10.	1995	Craig Venter, Hamilton Smith, and colleagues at The Institute for Genomic Research (TIGR) published the first complete genome of a free-living organism, the bacterium Haemophilus influenzae
11.	1996	Pal Nyren and his student Mostafa Ronaghi at the Royal Institute of Technology in Stockholm published their method of pyrosequencing
12.	1998	Phil Green and Brent Ewing of the University of Washington publish ‘‘phred’’ for sequencer data analysis
13.	2000	Lynx Therapeutics publishes and markets ‘‘MPSS’’—a parallelized, adapter/ligation-mediated, bead-based sequencing technology, launching ‘‘next-generation’’ sequencing
14.	2001	A draft sequence of the human genome published
15.	2004	454 Life Sciences markets a parallelized version of pyrosequencing. The first version of their machine reduced sequencing costs 6-fold compared to automated Sanger sequencing, and was the second of a new generation of sequencing technologies, after MPSS
16.	2005	Solexa/ Illumina sequence analyzer which gave an output data of 10E+7 Kbp
17.	2010	Illumina Hi-seq 2000 was introduced which gave an output of 10E+8 Kbp

Automated DNA Sequencing

A major advance in determining DNA sequence information occurred with the introduction of automated DNA sequencing machines. The automated sequencer is used to separate sequencing reaction products, detect and collect (via computer) the data from the reactions, and analyse the order of the bases to automatically deduce the base sequence of a DNA fragment. Automated sequencers detect extension products containing a fluorescent tag, allowing researchers to eliminate radioactivity from the DNA sequencing process. Sequence lengths that can be read using an automated sequencer

Published: 30 October 2024
ISSN: 2583-7850

are dependent upon a variety of parameters, but typically range between 500 and 1000 bases. As described for Sanger-type sequencing reactions using (primarily) isotopes to detect the extension products, some automated sequencers use four lanes to collect the data from the reactions. However, some machines use differently coloured fluorescent tags to indicate base identity (Hardin 2001). This approach enables a single lane to contain the data for a DNA template and increases fourfold the amount of data contained on a gel. This single-lane approach is made possible by the development of fluorescent tags that can be attached either to the DNA primer or to the ddNTP.

CONCLUSION

Many complex diseases, including cancer, have been diagnosed and treated using sequencing techniques. Although methods such as reverse transcription PCR (RT-PCR), fluorescence in-situ hybridization (FISH), and immunohistochemistry are used in the diagnosis of diseases, sequencing methods have a greater advantage. Sequencing methods can pave the way for new perspectives in disease diagnosis and treatment. Genetic variants that are currently unknown can be identified by sequencing. In order to find alternative treatment methods, changes in DNA, RNA, or protein levels can be examined by sequencing methods, mutations in cancer signaling pathways or repair mechanisms can be detected, and genetic profiles of individuals can be narrowed by comparing transcriptome readings of genetic materials obtained from patients and healthy individuals (Eren *et al.,* 2022). As a result, the disease can be diagnosed early on, paving the way for personalized treatment methods. Multiple sequencing methods have been introduced to date. These sequencing methods differ in terms of read speed, accuracy, and cost. Since DNA is a long material, the time, cost and required storage space of the sequencing method increase proportionally.

References

Behjati, S and Tarpey, P. S. 2013. What is next generation sequencing? *Arch Dis Child Educ Pract Ed*. 98(6):236-238.

Collins, F. S, Patrinos. A, Jordan. E, Chakravarti. A, Gesteland. R and Walters, L. R. 1998. New goals for the U.S. Human Genome Project: 1998-2003. *Science*. 282(5389):682-689.

Dogan M, Eroz. R, Yuce. H, Ozmerdivenli. R, Yeni Nesil Dizileme and Hakkinda Bilinenler. 2017. The Known about Next-Generation Sequencing (NGS). *Rev Literat.* 19(1):27-30.

Eren, K, Taktakoglu, N and I. Pirim. 2022. DNA Sequencing Methods: From Past to Present. *Eurasian J Med*.; 54(Suppl. 1): S47-S56.

Hardin, S. H. 2001. DNA Sequencing. *Encyclopedia of Life Sciences* Pp 1-5

Zhong Y, Xu F, Wu J, Schubert J and Li M. M. 2021. Application of next generation sequencing in laboratory medicine. *Ann Lab Med*. :41(1):25-43.

Published: 30 October 2024
ISSN: 2583-7850