# DNA Sequencing evolution

Nabil Youssef, STEM October High School

Abstract

*The genetic code is a universal language present in all known living organisms. The sequence of the four bases (adenine guanine thymine and cytosine) determines the genotype and phenotype of a living being. DNA sequencing can be used to determine the nucleotide sequence of specific genes, larger genetic regions, whole chromosomes, or the entire genome of an organism. Knowing this helps scientists answer fundamental biological questions about evolution and how life works. Known genomes in humans can be scanned for diseases and plants modified to create GM crops that are resistant to pests or have a higher yield. This technology is crucial to all genetic engineering. This article will cover the evolution of DNA Sequencing and explain the complete procedure of two of the most common methods of sequencing DNA, Sanger sequencing as well as next-generation sequencing.*

## I. Introduction

The last five to 10 years have seen some extraordinary feats in biology, among them determination of the entire DNA sequences of several extinct species, including woolly mammoths, Neanderthals, and a 700,000-year-old horse. Pivotal to those discoveries was the sequencing of the human genome, essentially completed in 2003 [1]. This endeavor marked a turning point in biology because it sparked remarkable technological advances in DNA sequencing. The primary human genome sequence took several years at a price of 1 billion dollars; the time and price of sequencing a genome are in free fall since then [1].

The discovery of the structure of the DNA molecule (Figure 1), with its two complementary strands, opened the door for the event of DNA sequencing and lots of other techniques utilized in scientific research today. Key to several of those techniques is macromolecule hybridization, the pairing of 1 strand of DNA to the complementary sequence on a strand from another DNA molecule.

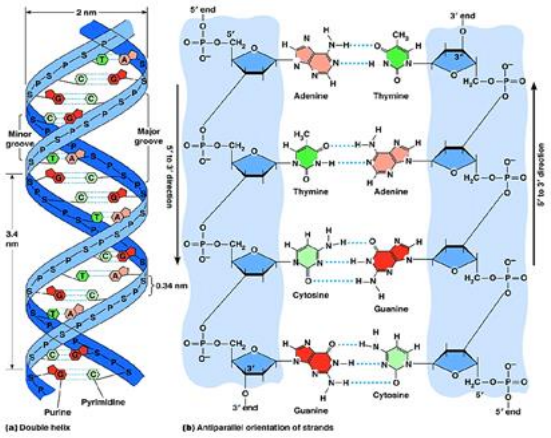



FIGURE 1 STRUCTURE OF THE DNA MOLECULE

## II. The invention of DNA Sequencing

Early attempts to sequence DNA were unwieldy. In 1968, Wu and Kaiser reported the utilization of primer extension methods to work out 12 bases of the cohesive ends of bacteriophage lambda [2]. In 1973, Gilbert and Maxam reported 24 bases of the lactose-repressor binding site, by copying it into RNA and sequencing those fragments. This took two years; one base per month [3].

In around 1976, the development of two methods that would decode many bases in a day transformed the sector [4]. Both methods, the chain terminator procedure developed by Sanger and Coulson, and the chemical cleavage procedure developed by Maxam and Gilbert, used distances along a DNA molecule from a radioactive label to positions occupied by each base to find out nucleotide order. Sanger's method involved four extensions of a labeled primer by DNA polymerase, each with trace amounts of 1 chain-terminating nucleotide, to supply fragments of various lengths [4]. Gilbert's method took a terminally labeled DNA-restriction fragment, and, in four reactions, used chemicals to make base-specific partial cleavages [4]. For both methods, the sizes of fragments present in each base-specific reaction were measured by electrophoresis on polyacrylamide slab gels, which enabled the separation of the DNA fragments by size with single-base resolution. The gels, with one lane per base, were put onto X-ray film, producing a ladder image from which the sequence might be read off immediately, rising the four lanes by size to infer the order of bases.

## III. Dideoxy Chain Termination Method for Sequencing DNA

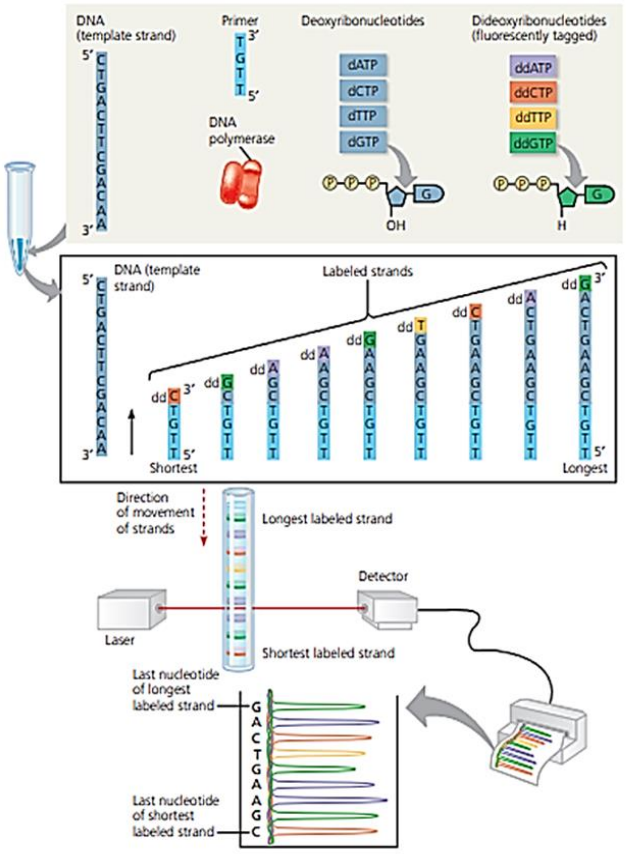
Sanger sequencing, also referred to as chain-termination sequencing, refers to a way of DNA sequencing developed by Sanger in 1977 (Figure 2). This method is predicated on the synthesis of a nested set of DNA strands complementary to one strand of a DNA fragment. Each new strand starts with an equivalent primer and ends with a dideoxy ribonucleotide (ddNTP), a modified deoxyribonucleotide (dNTP). The incorporation of a ddNTP terminates a growing DNA strand because it lacks a 3' OH group, the location for attachment of subsequent nucleotide. within the set of latest strands, each nucleotide position along the first sequence is represented by strands ending at that time with the complementary ddNTP.

**Procedure**

1) The fragment of DNA to be sequenced is denatured into single strands and incubated in a test tube with the necessary ingredients for DNA synthesis

2) Synthesis of each new strand starts at the 3′ end of the primer and continues until a ddNTP happens to be inserted instead of the equivalent dNTP. The incorporated ddNTP prevents further elongation of the strand. Eventually, a set of labeled strands of every possible length is generated, with the color of the tag representing the last nucleotide in the sequence.

3) The labeled strands in the mixture are separated by passage through a gel that allows shorter strands to move through more quickly than longer ones. For DNA sequencing, the gel is in a capillary tube, and its small diameter allows a fluorescence detector to sense the color of each fluorescent tag as the strands




come through. Strands differing in length by as little as one nucleotide can be distinguished from each other.

Because each type of ddNTP is tagged with a unique fluorescent label, the identity of the ending

nucleotides of the new strands, and ultimately the entire original sequence, can be determined. The color of the fluorescent tag on each strand indicates the identity of the nucleotide at its 3'-end. The results can be printed out as a spectrogram.

## IV. Next-Generation Sequencing

The main difference between NGS (Figure 3) and Sanger sequencing is the construction of the sequencing library. Sanger sequencing libraries need multiple steps that combine molecular biology with microbiological culture to represent the DNA sample of interest as a series of subclones in a bacterial plasmid or phage vector. These subclones then need growth in culture and DNA isolation before sequencing. This multistep process can be completed in approximately one week, at which point the purified DNAs are ready for sequencing [3]. On the other hand, the simplicity and speed of NGS library construction are remarkable. Starting from a variety of input DNA sources ranging from high molecular weight genomic DNA to a pool of PCR products, to short stretches of histone-bound DNA released after chromatin immunoprecipitation (ChIP) or reverse transcriptase–converted RNA [1].

**Procedure**

1) Genomic DNA is fragmented, and fragments of 400 to 1,000 base pairs are selected.

2) Each fragment is isolated with a bead in a droplet of aqueous solution.

3) The fragment is copied over and over by a technique called PCR (to be described later). All the 5′ ends of one strand are specifically" captured" by the bead. Eventually, 106 identical copies of the same single strand, which will be used as a template strand, are attached to the bead.

4) The bead is placed into a small well along with DNA polymerases and primers that can hybridize to the 3′ end of the single (template) strand.

5) The well is one of 2 million on a multiwell plate, each containing a different DNA fragment to be sequenced. A solution of one of the four nucleotides is added to all wells and then washed off. This is done sequentially for all four nucleotides: dATP, dTTP, dGTP, and then dCTP. The entire process is then repeated.

6) In each well, if the next base on the template strand (T in this example) is complementary to the added nucleotide (A, here), the nucleotide is joined to the growing strand, releasing PPi, which causes a flash of light that is recorded.

7) The nucleotide is washed off and a different nucleotide (dTTP, here) is added. If the nucleotide is not complementary to the next template base (G, here), it is not joined to the strand and there is no flash.

8) The process of adding and washing off the four nucleotides is repeated until every fragment has a complete complementary strand. The pattern of flashes reveals the sequence of the original fragment in each well.

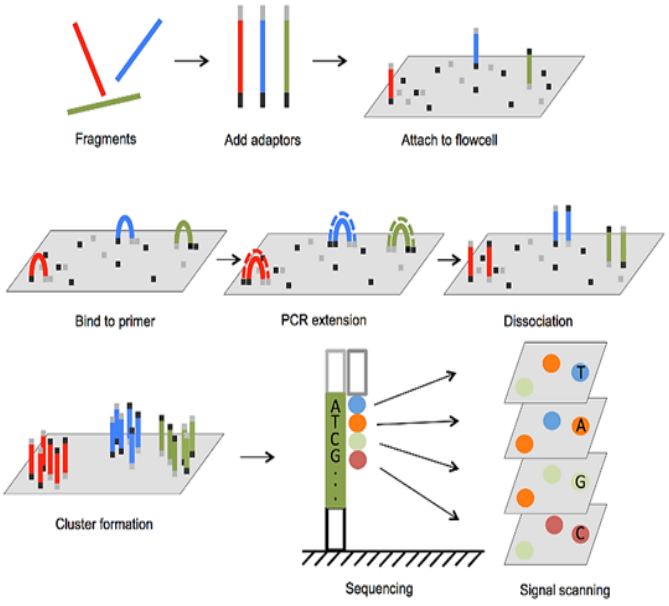



FIGURE 3 NEXT-GENERATION SEQUENCING

## V. Conclusion

Improved DNA sequencing techniques have transformed the way in which we can explore fundamental biological questions about evolution and how life works. Little more than a decade after the human genome sequence was announced, researchers had completed sequencing roughly 4,000

bacterial, 190 archaeal, and 180 eukaryotic genomes, with more than 17,000 additional species underway [1]. Complete genome sequences have been determined for cells from several cancers, for ancient humans, and for the many bacteria that live in the human intestine.

## VI. References


[1] J. Shendure et al., "DNA sequencing at 40: past, present and future", 2021.

[2] Wu, R. & Kaiser, A. D. Structure, and base sequence in the cohesive ends of bacteriophage lambda DNA. J. Mol.

[3] Gilbert, W. & Maxam, A. The nucleotide sequence of the lac operator.

[4]. Sanger, F., Nicklen, S. & Coulson, A. R. DNA sequencing with chain-terminating inhibitors.