THE SCIENCE WORLD
A Monthly e Magazine
ISSN:2583-2212
Extended Popular Article
Sept, 2023; 3(09), 2378-2393

DNA Sequencing techniques: Sanger to Next Generation Sequencing

Manoj Kumar Saini[1], Gaurav[2], Harish BM[3], Jitendra Kumar [3] and Kumar Sanu[4]
[1]Ph.D. Research Scholar Department of Genetics and Plant Breeding, Chaudhary Sarwan Kumar Himachal Pradesh Krishi Vishvavidyalaya, Palampur-176062 (HP), India
[2]Ph.D. Research Scholar Department of Biotechnology, Chaudhary Sarwan Kumar Himachal Pradesh Krishi Vishvavidyalaya, Palampur-176062 (HP), India
[3]Ph.D. Research Scholar Department of Vegetable Science and Floriculture, Chaudhary Sarwan Kumar Himachal Pradesh Krishi Vishvavidyalaya, Palampur-176062 (HP), India
[4]M.Sc. Scholar Department of Genetics and Plant Breeding, Chaudhary Sarwan Kumar Himachal Pradesh Krishi Vishvavidyalaya, Palampur-176062 (HP), India
https://doi.org/10.5281/zenodo.8376905

Abstract

Numerous research applications require that the order of nucleic acid residues in biological samples be determined. Numerous scientists have worked hard over the past 50 years to develop the methods and tools necessary to accomplish this feat of sequencing DNA and RNA molecules. The sequencing of small oligonucleotides has given way to sequencing millions of bases and the struggle to deduce the coding sequence of a single gene has given way to whole genome sequencing that is quick and widely accessible. The sequence of the nucleotides in a DNA molecule is determined via sequencing technology. Each creature has a certain order that determines who we are, how we act, and how we respond to a changing environment. Next Generation Sequencing and Sanger are the two primary methods.

Introduction

"Knowledge of sequences could contribute much to our understanding of living matter."

Frederick Sanger

Deciphering a species' unique DNA sequence therefore entails comprehending the molecular underpinning of life. The polynucleotide chains' nucleic acid sequence ultimately determines the genetic and metabolic characteristics of terrestrial life. Therefore, it is crucial for biological research to be able to measure or infer such sequences. There are two primary methods of sequencing, Sanger sequencing and Next-Generation Sequencing (NGS). Long-read sequencing and short-read sequencing are two further NGS subsets.

Official Website
www.thescienceworld.net
thescienceworldmagazine@gmail.com
Published: **24.09.2023**

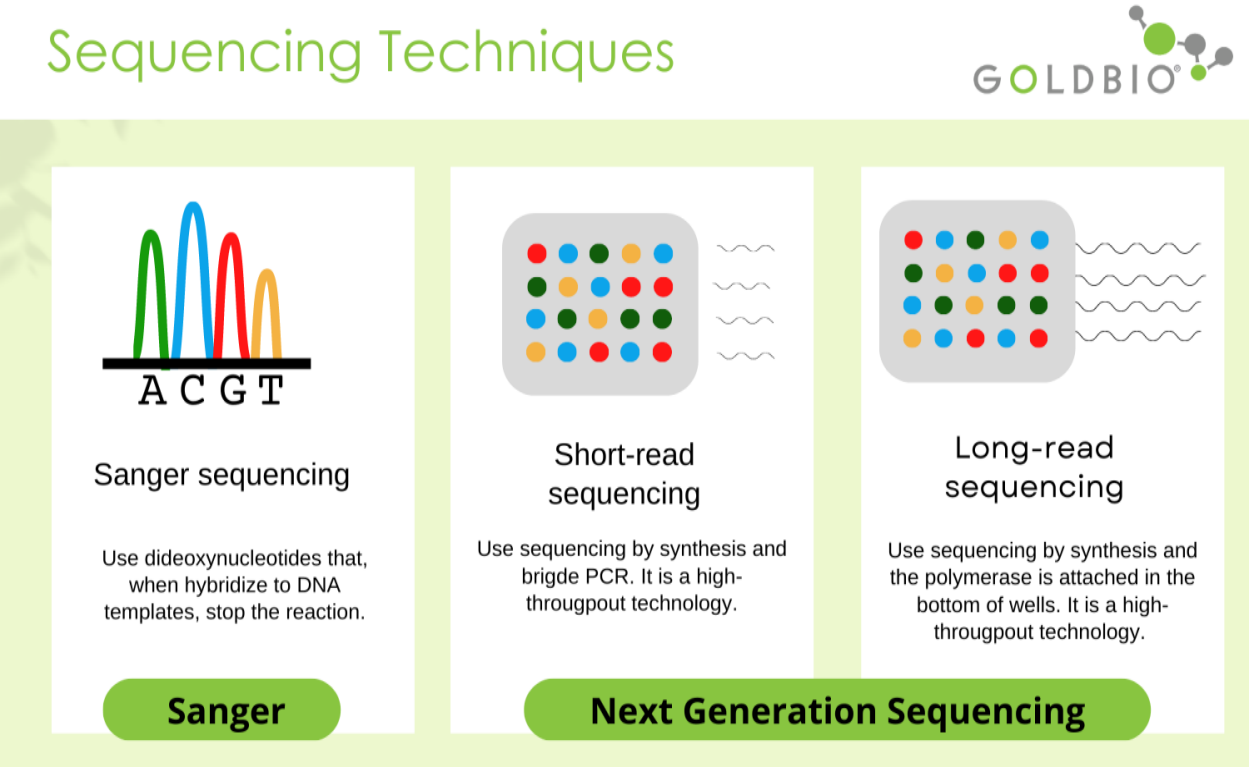


Figure 1. Illustrates the various sequencing strategies.

The generations it is classified into include:

First Generation

- Sanger Sequencing

Second Generation Sequencing

- Pyrosequencing
- Sequencing by Reversible Terminator Chemistry
- Sequencing by Ligation

Third Generation Sequencing

- Single Molecule Fluorescent Sequencing
- Single Molecule Real Time Sequencing
- Semiconductor Sequencing
- Nanopore Sequencing

Fourth Generation Sequencing

Aims conducting genomic analysis directly in the cell.

Sanger sequencing

Chain termination is another name for Sanger sequencing. This works by giving the DNA polymerase radiolabeled dideoxy nucleotides as substrates. In contrast to typical DNA nucleotides, dideoxy nucleotides have a H group rather than an OH group in the sugar molecule. This is essential for the Sanger sequencing method since the reaction comes to a halt once the polymerase uses one dideoxy nucleotide to hybridize with the DNA template. Or, to put it another way, the chain breaks.

Official Website
www.thescienceworld.net
thescienceworldmagazine@gmail.com
Published: 24.09.2023

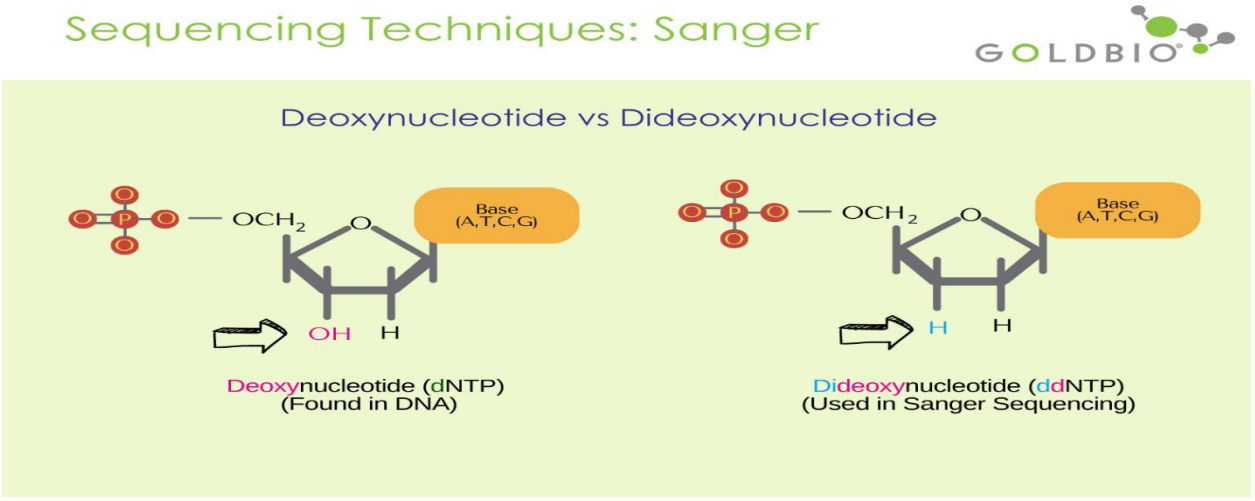


Figure 2. Differences between dideoxy nucleotide and deoxynucleotide

The order of the nucleotides in a DNA template was discovered by Sanger and colleagues using this fundamental idea. Sanger took over four years to decrypt the 5000 bases of the virus phiX174 using his own method! Three steps were employed in traditional Sanger sequencing. Making four separate reactions in various tubes was the first stage. A primer, dNTPs (standard nucleotides), DNA polymerase, and a little quantity of a particular radiolabeled dideoxy nucleotide were all mixed together in each tube. Adenine (ddATP), thymine (ddTTP), cytosine (ddCTP), and guanine (ddGTP) are examples of the dideoxy nucleotide (ddTP).

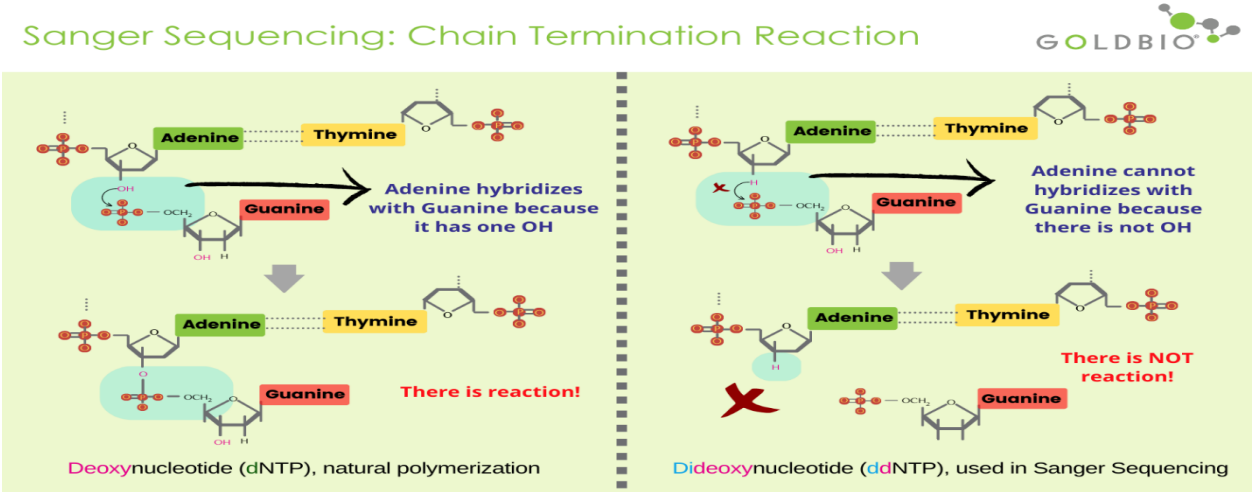


Figure 3. Chain termination reaction in Sanger sequencing.

Let's take ddATP as an example and look inside a single tube. Here, you'll find a number of varied length sequences that all terminate in the same radiolabeled ddATP. The position of adenines in the sequence can then be approximated if the size of the DNA fragment is known. Autoradiography and polyacrylamide gel electrophoresis were utilized in the second step. For each tube, this step displayed a different band. The third and last phase was determining the nucleotide order. Sanger sequencing, however, developed into a contemporary method. Modern sequencing eliminates the need for four separate reactions and allows for one reaction to be performed in a single tube by attaching a distinct fluorescent dye to each type of dideoxy nucleotide. Then, contemporary sequencing use microfluidics

Official Website
www.thescienceworld.net
thescienceworldmagazine@gmail.com
Published: 24.09.2023

and potent computers to detect the nucleotide being sequenced in place of gels.

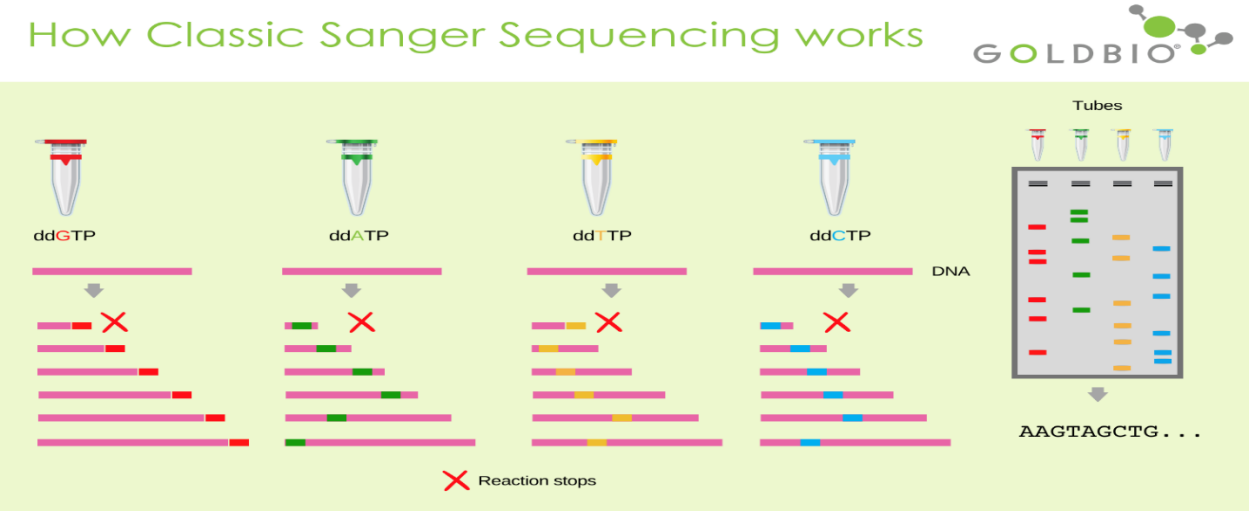


Figure 4. Conventional Sanger sequencing process.

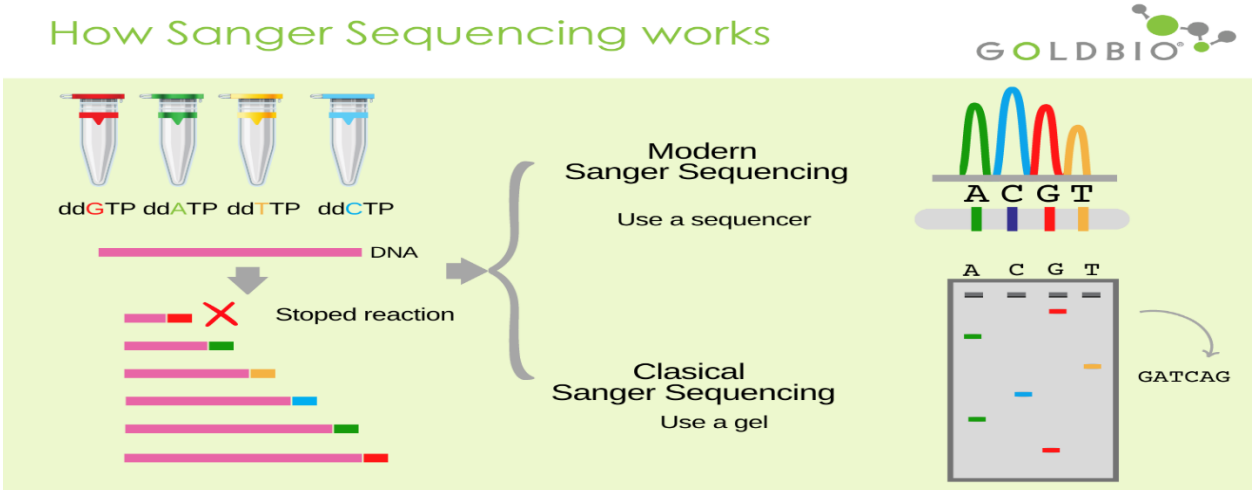


Figure 5. Overview Sanger sequencing.

Next-Generation DNA Sequencing

Sequence-based procedures, as opposed to microarray techniques, directly identify the nucleic acid sequence of a specific DNA or cDNA molecule. The Human Genome Project was the first significant attempt at DNA sequencing. Sanger sequencing, also known as the chain-termination method, was utilized in this research, which took 13 years, cost $3 billion, and was finished in 2003. The short read, massively parallel sequencing technique is fundamentally different from traditional Sanger sequencing using capillary electrophoresis. It revolutionized sequencing capabilities and gave rise to next-generation sequencing (NGS), which offers orders of magnitude more data at a significantly lower recurring cost.

The phrase "next-generation sequencing" (NGS), commonly referred to as "high-throughput sequencing," is used to refer to a variety of contemporary sequencing technologies. The study of genomics and molecular biology has been completely transformed by these technologies, which make it possible to sequence DNA and RNA considerably more swiftly and affordably than was previously possible using Sanger sequencing. These technologies consist of:

Official Website
www.thescienceworld.net
thescienceworldmagazine@gmail.com
Published: 24.09.2023

Advantages of NGS

NGS is a common technology in functional genomics and can be used to analyze DNA and RNA materials. NGS-based approaches provide a number of advantages over microarray techniques, including:

- It gives single-nucleotide resolution, making it possible to find related genes (or features), alternatively spliced transcripts, allelic gene variants, and single nucleotide polymorphisms.
- It does not require a priori knowledge of the genome or genomic features.
- Higher dynamic range of signal.
- Less input DNA/RNA required (nanograms of materials are adequate).
- Higher reproducibility.

Next Generation Sequencing (NGS) is a robust platform that has enabled the sequencing of thousands to millions of DNA molecules simultaneously.

NGS: Short-read Sequencing (SRS)

Due to its high throughput capabilities, short-read sequencing (SRS), spearheaded by Illumina, is currently the most used sequencing method. Second Next Generation Sequencing is another name for short-read sequencing. SRS employs the chemical process known as sequencing by synthesis. Additionally, thousands of data are produced because of the numerous reactions that take place during the clustering process even though SRS also makes use of various fluorophores added to each nucleotide (similar to contemporary Sanger sequencing). Hundreds of DNA molecules are combined with fluorescent dideoxy nucleotides (ddTPs) and connected to a flow cell (a slide with numerous small lanes) to form clusters. A high-resolution computer then records the fluorescence of millions of simultaneous reactions. It is significant to note that, in contrast to sanger, NGS uses millions of sequences. In NGS, reassembling short sequences based on overlapping portions is the aim. Imagine that you need to read a book, but the pages have been torn, and you need to put them back together correctly. How does one create it? Comparable to NGS, overlapping words are utilized as hints to correctly reassemble the genome. The drawback of this is that, despite the employment of robust algorithms to automate the reassembling, they are not always precise. Better algorithms and perhaps even software powered by artificial intelligence will most certainly be developed in the upcoming years to assist us in improving the reconstruction of genomes and transcriptomes.

SRS can be carried out utilizing Illumina technology in two basic steps.

The clustering stage comes first, following the preparation of the library. An isothermal technique is used to amplify DNA molecules with attached adapters (temperature remains constant but volume and pressure may change). The DNA templates being sequenced can be amplified using this technique. **Thousands of oligos, or DNA short sequences, are present in each line of the flow cell and bind to the adapters at the ends of the DNA molecules by complementation. Some oligos**

Official Website
www.thescienceworld.net
thescienceworldmagazine@gmail.com
Published: 24.09.2023

attach to the 3' and 5' extremities, whereas other oligos bind to both. Through a procedure known as bridge clustering or bridge PCR, the oligos amplify DNA templates. One side of the DNA attaches to the first kind of oligo in the flow cell during the initial step. DNA polymerases and nucleotides that hybridize with the DNA template are present in the process. The second stage involves the folding over of DNA templates and the hybridization of the second type of oligo to the other end of the DNA molecule. The polymerization process then repeats. Therefore, DNA molecules are amplified and cloned at both ends. Thousands of DNA molecules are created after clustering and are connected to the flow cell.

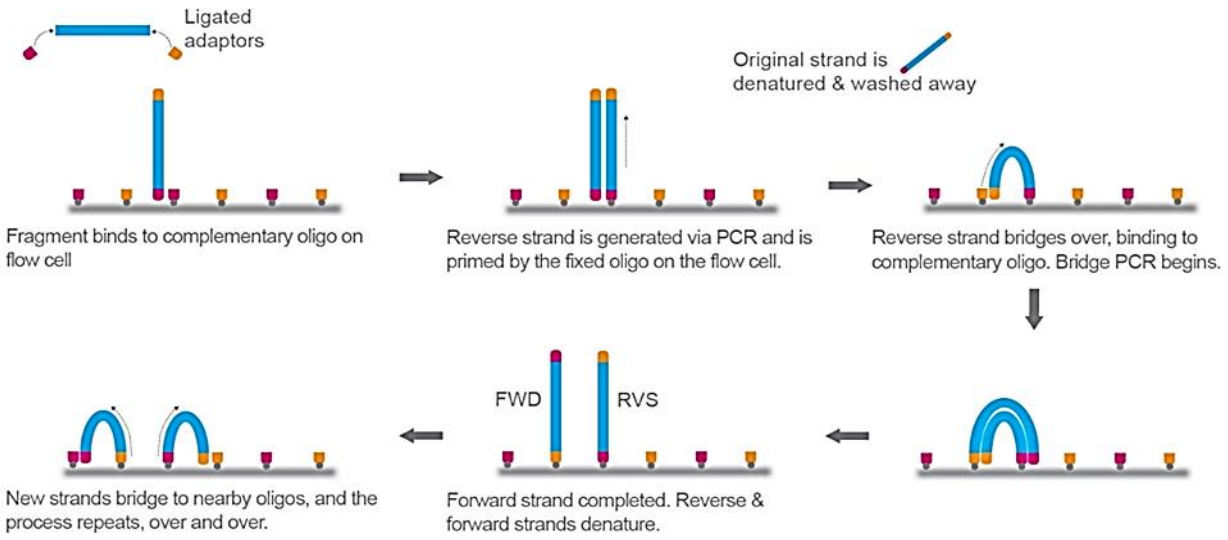


Figure 6. Bridge PCR used in short-read sequencing.

NGS Long-read DNA sequencing (LRS)

A contemporary technique called long-read DNA sequencing (LRS) is utilized to generate long and comprehensive DNA sequences. Third Next Generation Sequencing is another name for it. Although bridge amplification is not done, it uses the same sequencing by synthesis chemistry as short-read sequencing. Long-read sequencing, which differs from short-read sequencing in that it generates huge sequences from a single DNA template, has a larger error rate (defined as a lack of accuracy in the correct nucleotide order) than short-read sequencing (up to a 10% mistake). Long-read sequencing is the best method for determining the sequence of complex DNA regions like repeats (a region made up of numerous nearby copies of the same sequence), despite having a greater error rate in base calling. In cereals species where repeat regions are common; their genomes can include up to 80% repeats (Wicker *et al.*, 2001). The sequencing by synthesis process takes place in the second stage. Fluorescent nucleotides and DNA polymerases are combined in the flow cell in this instance, where they quickly hybridize with the many DNA molecules. A potent computer simultaneously records the fluorescence given out by the various processes to create millions of tiny simulated sequences known as reads.

Official Website
www.thescienceworld.net
thescienceworldmagazine@gmail.com
Published: 24.09.2023

There are two main phases involved in long-read sequencing. In the first, a well plate with only 100 nm in height has thousands of DNA templates primed to polymerases. Each well has a fixed DNA polymerase in the bottom as well as a small camera.

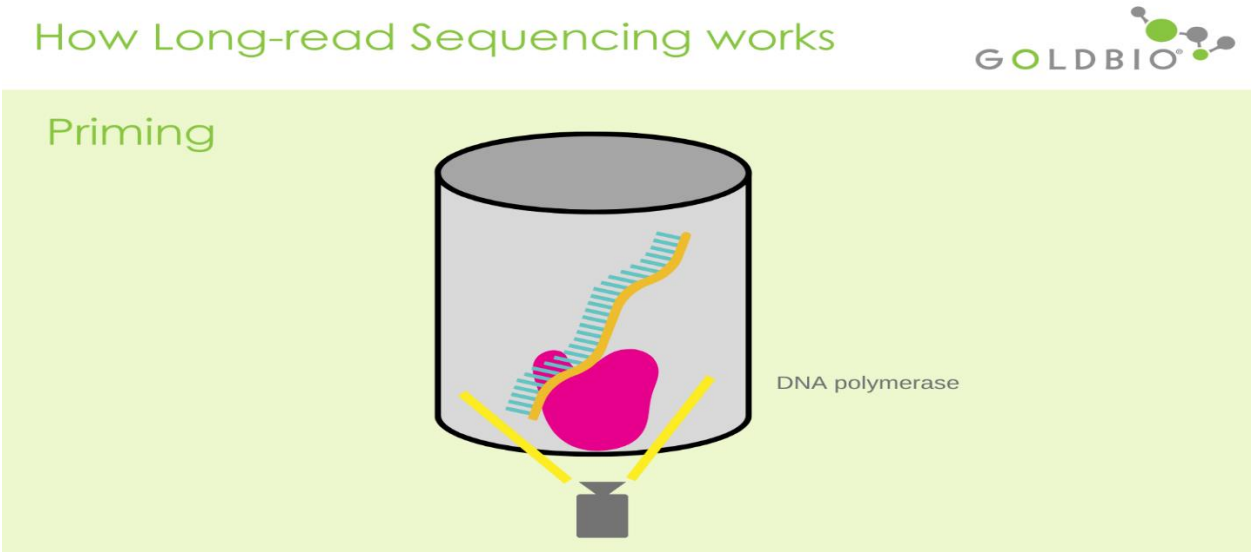


Figure 7. Priming process in long-read sequencing.

The second phase involves the ordered hybridization of fluorescent nucleotides with DNA templates, and the tiny camera records the fluorescence reactions. The signal becomes stronger and reveals the correct arrangement of the nucleotides when the right base attaches to the DNA template.

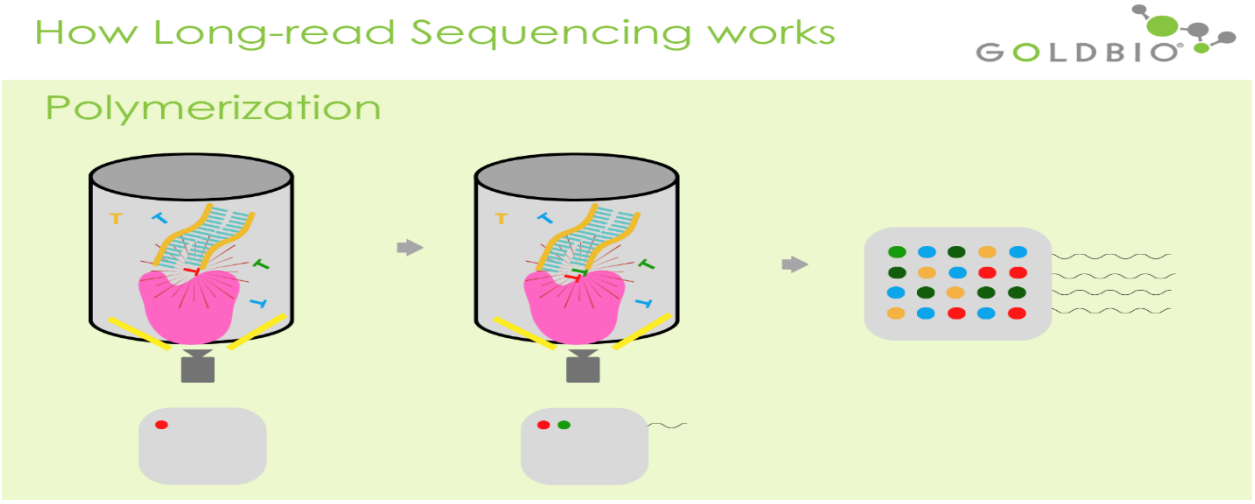


Figure 8. Hybridization of fluorescent nucleotides with DNA templates

Applications of NGS

The kind of questions that scientists can pose and receive answers to have been radically altered by next-generation sequencing technology. A wide range of applications are made possible by innovative sample preparation and data analysis techniques. For instance, NGS permits labs to:

- Utilize RNA sequencing (RNA-Seq) to find novel RNA variants and splice sites, or quantify mRNAs for gene expression studies.
- Rapidly sequence complete genomes. Deeply sequence target regions.

Official Website
www.thescienceworld.net
thescienceworldmagazine@gmail.com
Published: 24.09.2023

- Examine epigenetic elements such DNA-protein interactions and genome-wide DNA methylation.
- Sequence cancer samples to learn more about tumour subclones, uncommon somatic variations, and other topics.
- Research the human microbiome.
- Find new pathogens.

Lynx therapeutics' massively parallel signature sequencing (MPSS)

It's regarded as the pioneering "next-generation" sequencing technology. At Lynx Therapeutics, a business established in 1992 by Sydney Brenner and Sam Eletr, MPSS was created in the 1990s. MPSS is a technology for extremely high throughput sequencing. When used to create an expression profile, it makes nearly every transcript in the sample visible and calculates its precise expression level. The MPSS bead-based method was prone to sequence-specific bias or loss of particular sequences since it used a complicated method that involved adapter ligation and adapter decoding while reading the sequence in increments of four nucleotides. However, the MPSS output's key characteristics, which included millions of small DNA sequences, were indicative of later "next-gen" data formats. For MPSS, these were typically applied to the sequencing of cDNA in order to quantify the levels of gene expression.

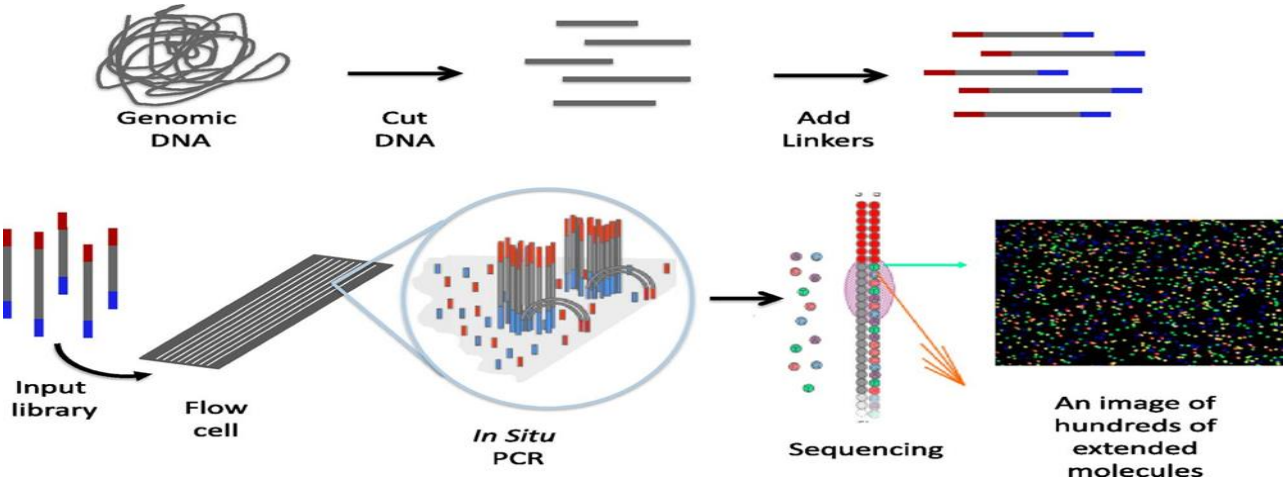


Figure 9. Lynx therapeutics' massively parallel signature sequencing (MPSS)

Polony sequencing

It is a cheap but very precise multiplex sequencing method that enables parallel reading of millions of immobilized DNA sequences. At Harvard Medical College, Dr. George Church invented this method first. It sequenced an E. coli genome with > 99.9999% accuracy at a cost of around 1/10 that of Sanger sequencing by using an in vitro paired-tag library, emulsion PCR, an automated microscope, and ligation-based sequencing chemistry.

Official Website
www.thescienceworld.net
thescienceworldmagazine@gmail.com
Published: 24.09.2023

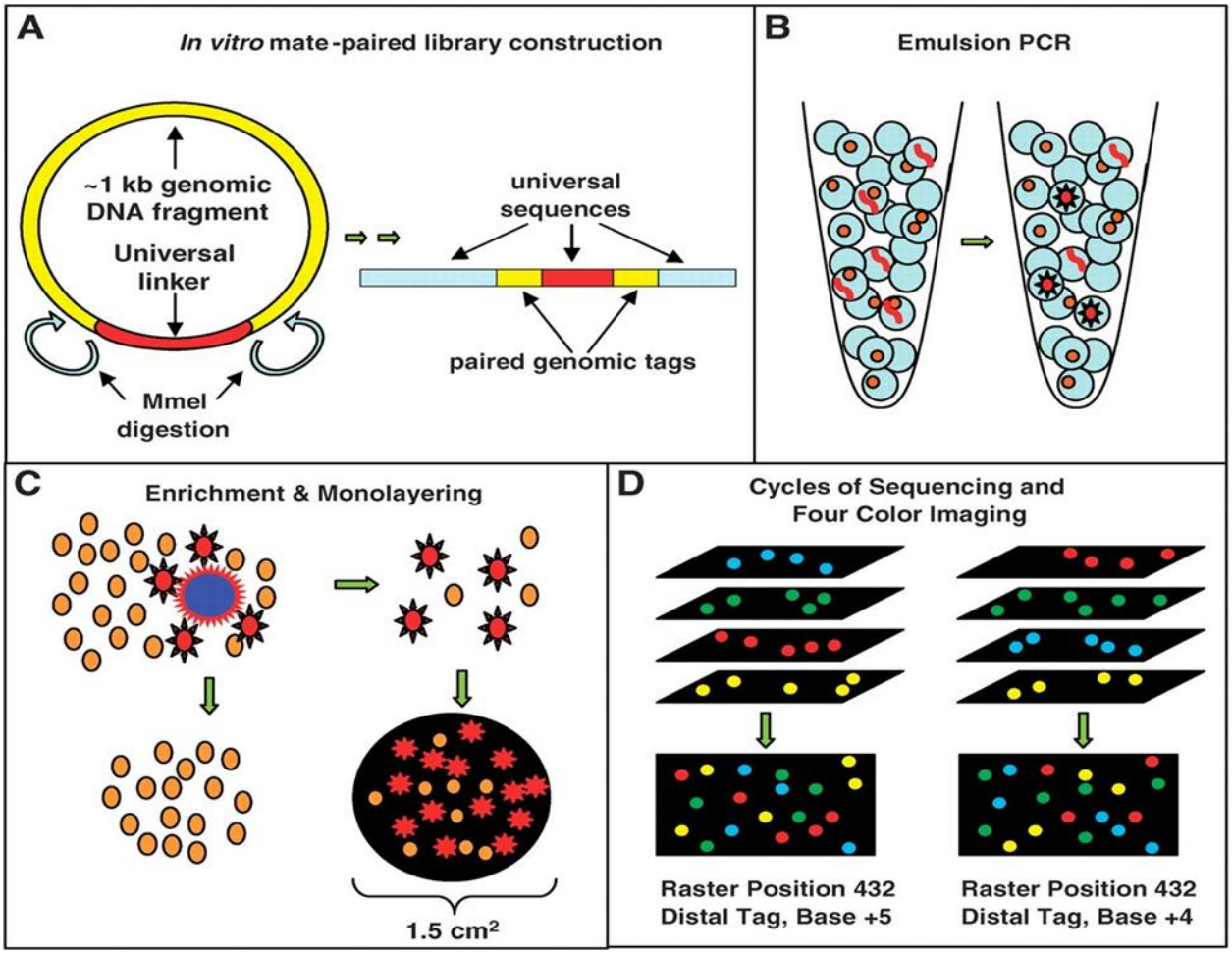


Figure 10. Polony sequencing

Pyrosequencing

Pyrosequencing was developed in parallel by 454 Life Sciences, which was later purchased by Roche Diagnostics. The technique, known as "emulsion PCR," amplifies DNA within water droplets suspended in an oil solution. Each droplet contains a single DNA template coupled to a single primer-coated bead, which subsequently establishes a clonal colony. There are several picolitre-volume wells in the sequencing machine, each housing a single bead and sequencing enzymes. Pyrosequencing makes use of luciferase to produce light for the purpose of detecting the individual nucleotides that have been added to the developing DNA. The resulting data are then merged to provide sequence read-outs. In comparison to Solexa and SOLiD on the other end and Sanger sequencing on the other, this method offers intermediate read length and cost per base.

Official Website
www.thescienceworld.net
thescienceworldmagazine@gmail.com
Published: 24.09.2023

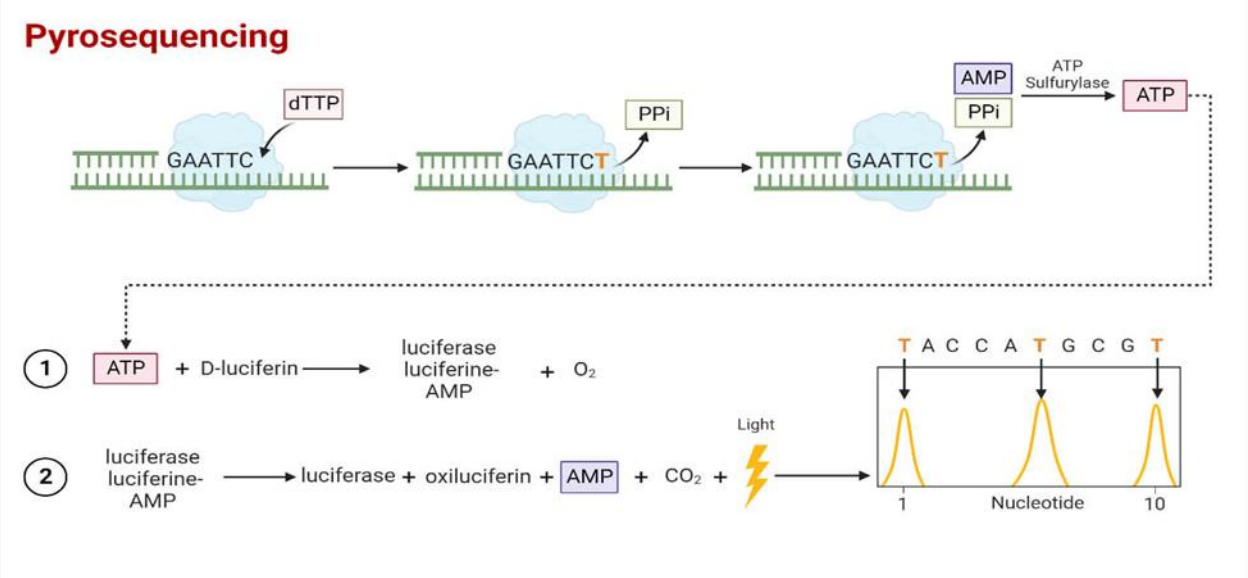
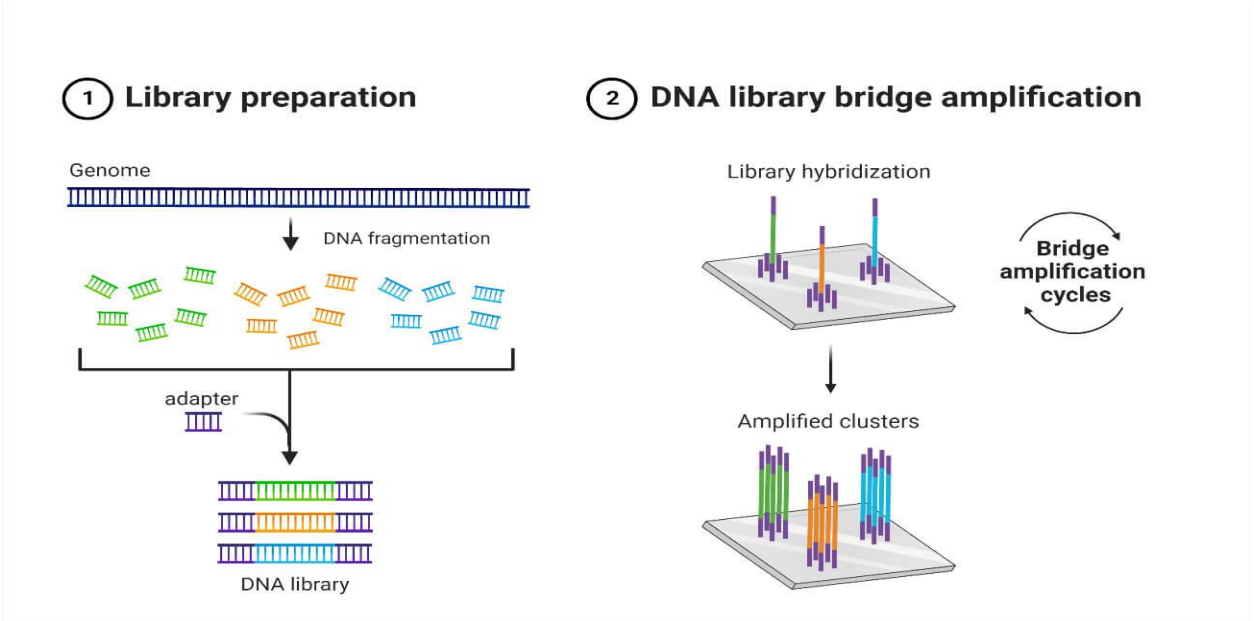


Figure 11. Pyrosequencing

Illumina (Solexa) sequencing

Solexa created a dye terminator-based sequencing method. In this technique, DNA molecules are initially amplified after being linked to primers on a slide. Bridge amplification is the term used for this. The DNA can only be stretched one nucleotide at a time, unlike pyrosequencing. The terminal 3′ blocker and dye are chemically removed from the DNA to start the next cycle after a camera captures photos of the fluorescently labelled nucleotides. Solexa created a dye terminator-based sequencing method. In this technique, DNA molecules are initially amplified after being linked to primers on a slide. Bridge amplification is the term used for this. The DNA can only be stretched one nucleotide at a time, unlike pyrosequencing. The fluorescently labelled nucleotides are photographed by a camera before the dye and terminal 3′ blocker is chemically removed from the DNA to start the



next cycle.

Official Website
www.thescienceworld.net
thescienceworldmagazine@gmail.com
Published: 24.09.2023

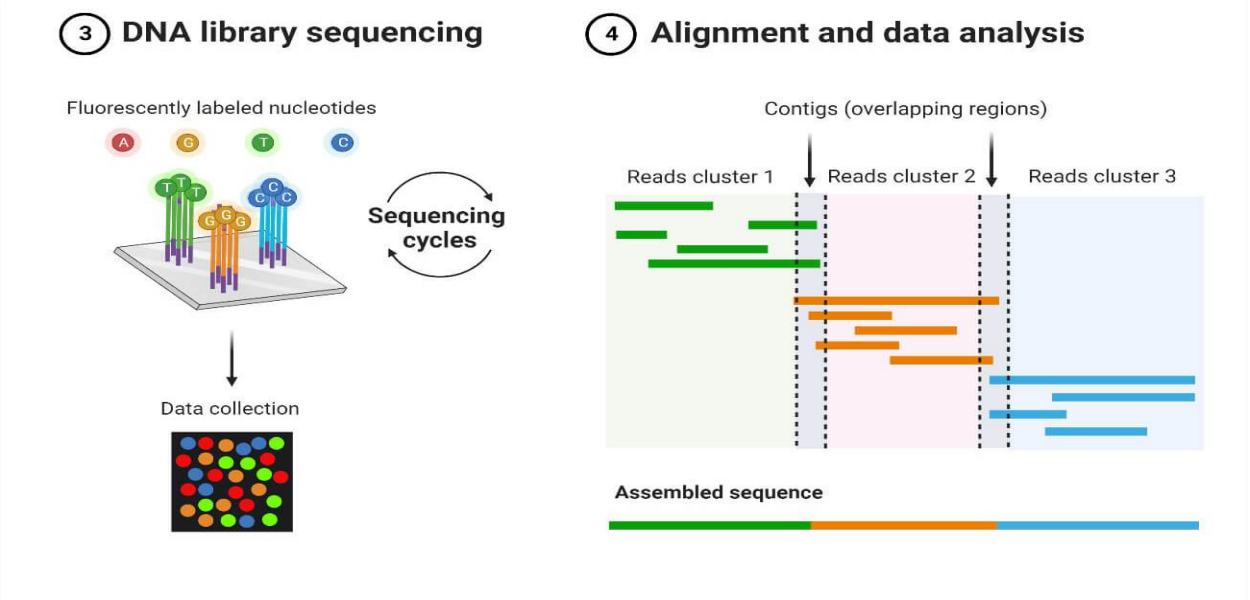
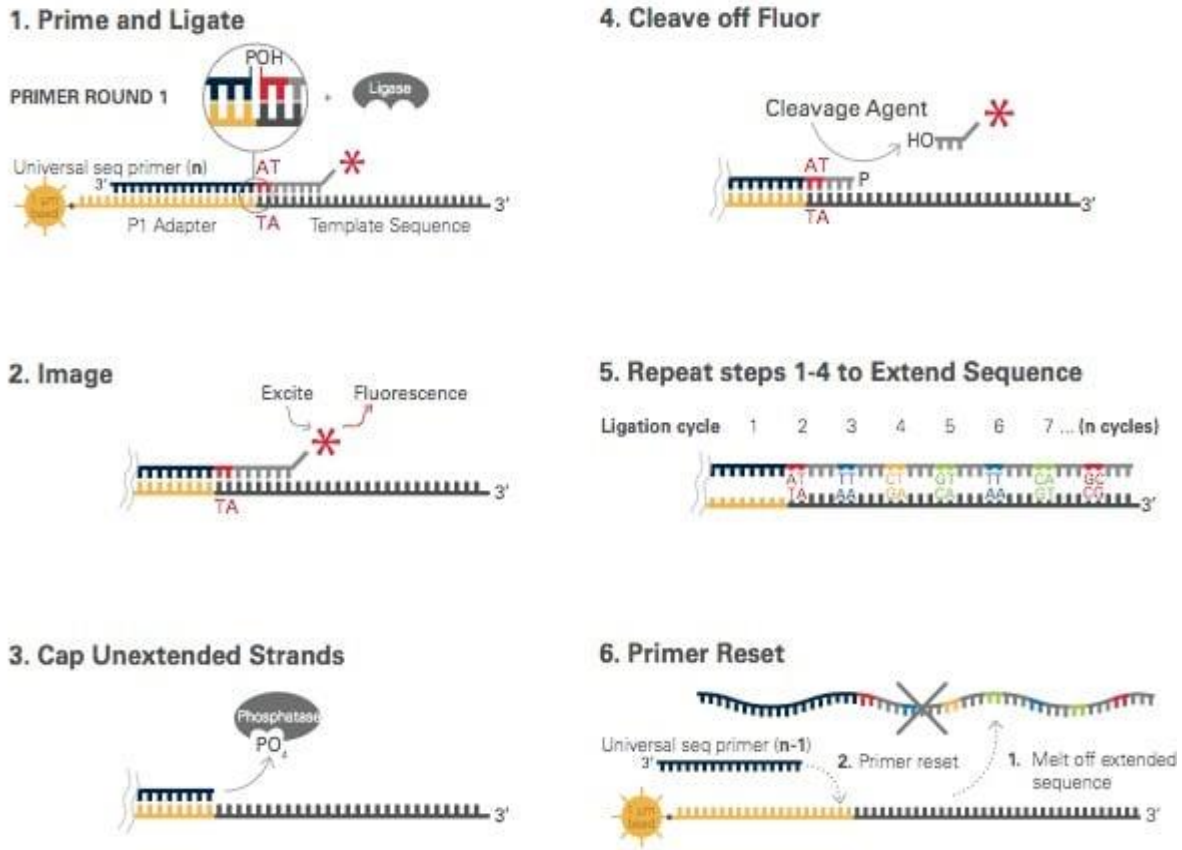


Figure 12. Illumina (Solexa) sequencing

SOLiD sequencing

Oligonucleotide ligation and detection is the sequencing method used in ABISolid sequencing.This involves labelling a pool of all potential oligonucleotides of fixed length in accordance with the sequenced position.Sequences produced by this sequencing are similar to illuminate sequencing in terms of quantity and length.



Official Website
www.thescienceworld.net
thescienceworldmagazine@gmail.com
Published: 24.09.2023

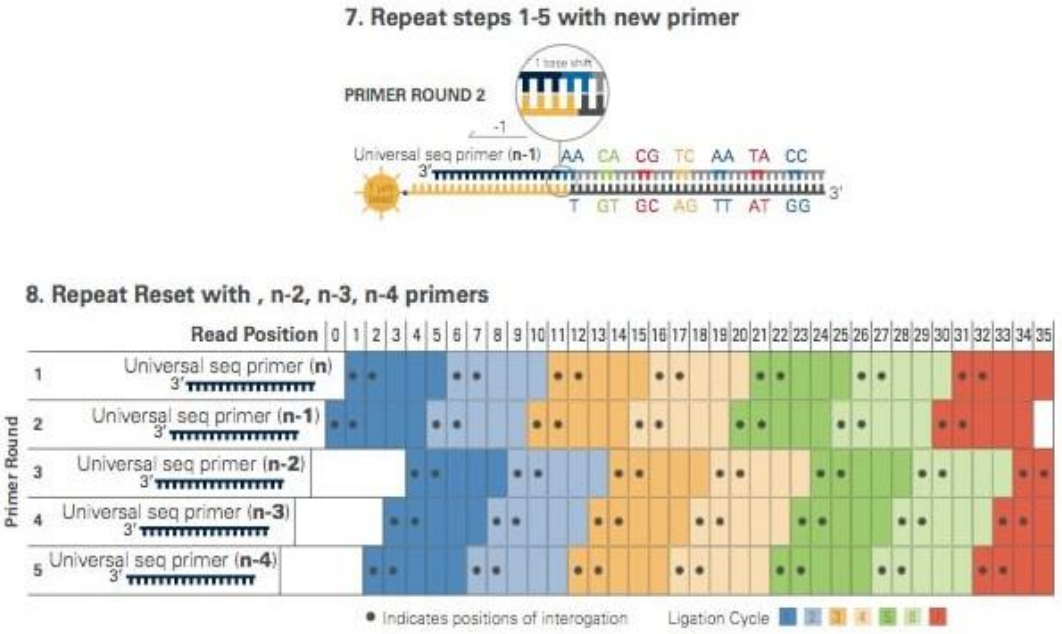


Figure 13. SOLiD sequencing

DNA Nanoball Sequencing

The full genetic sequence of an organism can be found using high throughput sequencing technologies. The technique amplifies genomic DNA molecule fragments via rolling circle replication. Compared to other next-generation sequencing platforms, this DNA sequencing method allows for the sequencing of a lot of DNA nanoballs in a single run at a low cost of reagents, but

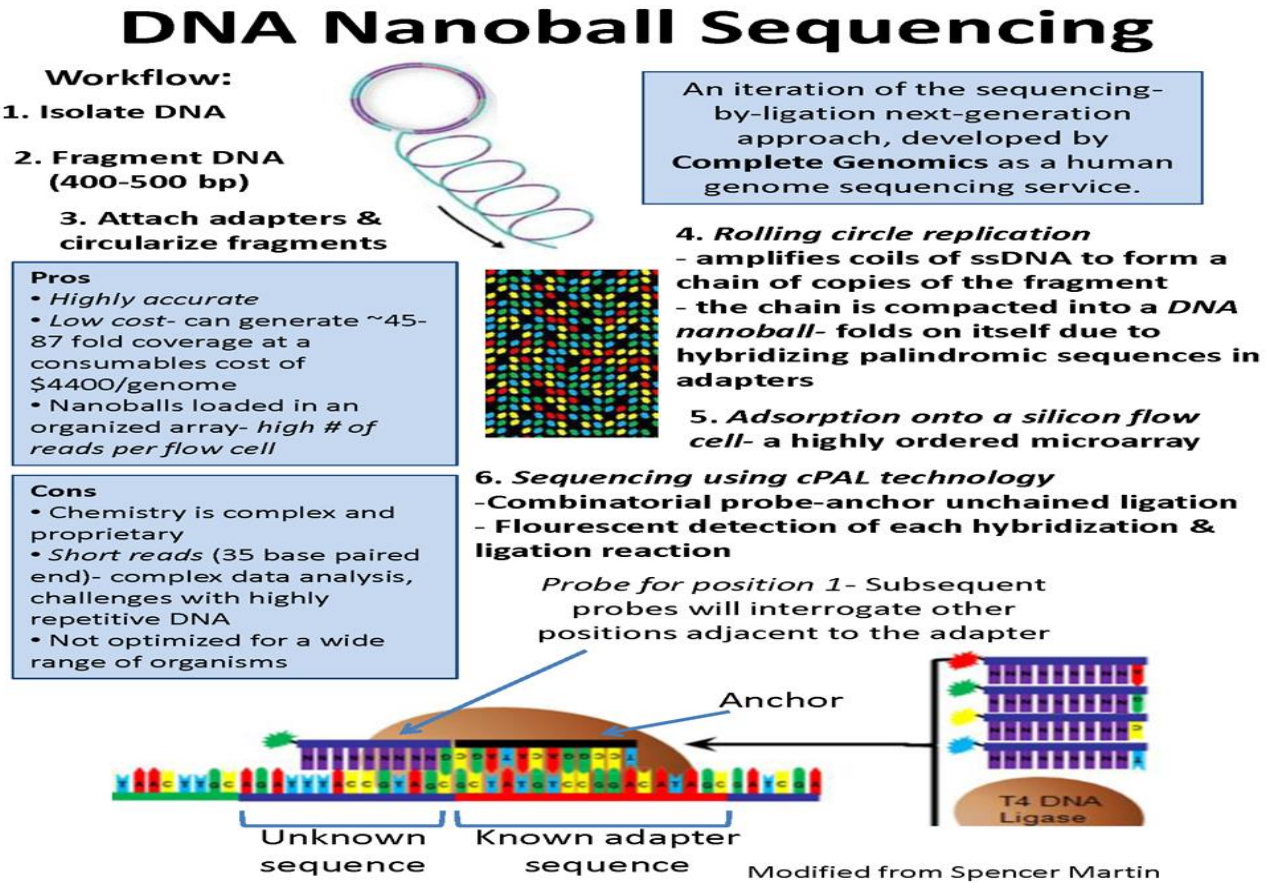


Figure 14. DNA Nanoball Sequencing

Official Website
www.thescienceworld.net
thescienceworldmagazine@gmail.com
Published: 24.09.2023

because each DNA nanoball can only yield short DNA sequences, it is challenging to map the short reads to a reference genome. Several genome sequencing initiatives have already used this method, and others are planned.

Helioscope single molecule sequencing

Helioscope sequencing employs DNA fragments attached to the flow cell surface with polyA tail adapters added. The following steps involve cyclic washing of the flow cell with fluorescently labelled nucleotides and extension-based sequencing. The Helioscope sequencer generates the reads, which are brief (up to 55 bases each run), but subsequent advances in the methodology have enabled more precise reads of homopolymers and RNA sequencing.

Single molecule SMRT sequencing

The sequencing by synthesis method is the foundation of SMRT sequencing. The so-called zero-mode wave-guides (ZMWs), which are little well-like containers with the capture instruments at the bottom, are where the DNA is synthesised. Unmodified polymerase and fluorescently tagged nucleotides that are freely dispersed in the solution are used to perform the sequencing. The wells are designed such that only the fluorescence present at the well's bottom may be seen. When the nucleotide is incorporated into the DNA strand, the fluorescent label separates from the nucleotide, leaving an unaltered DNA strand. The SMTR technology enables nucleotide changes to be found. This is accomplished by watching the polymerase kinetics. This method permits readings of 1,000 nucleotides.

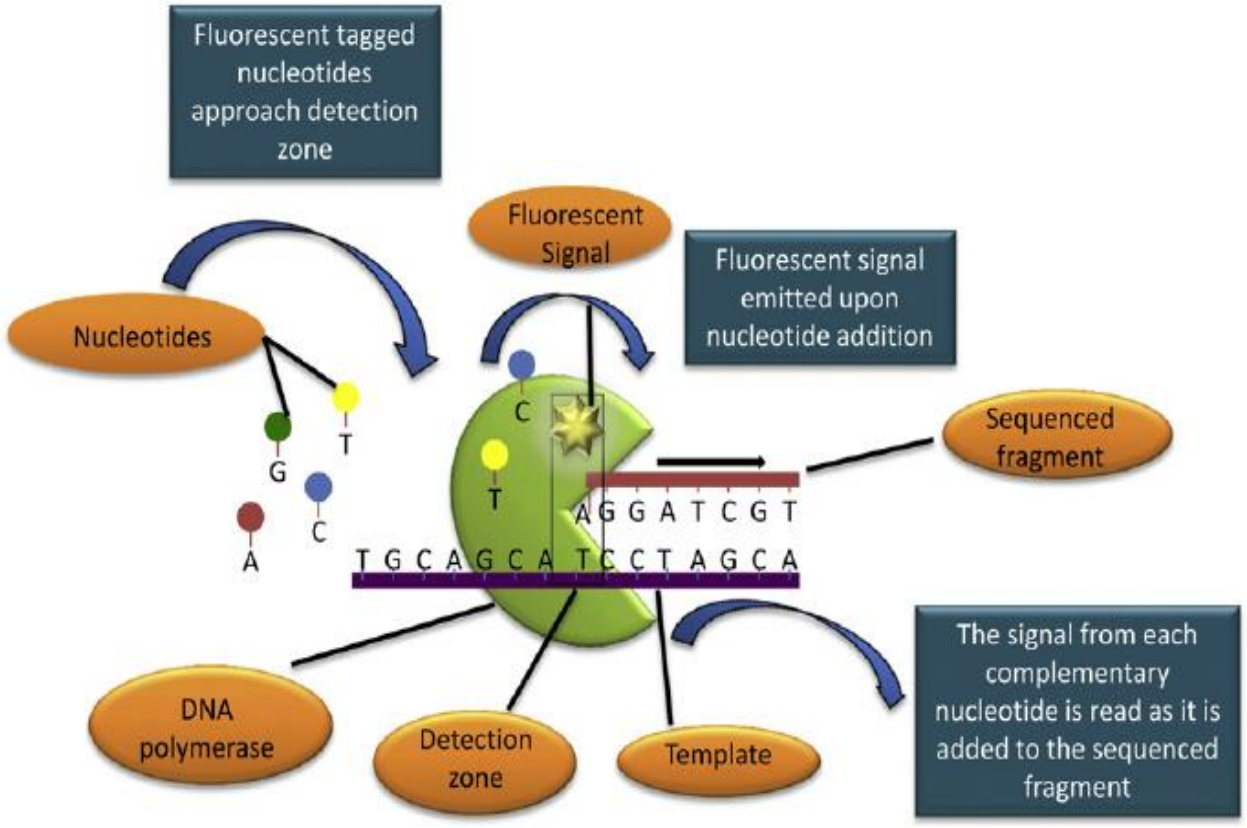


Figure 15. Single molecule SMRT sequencing

Official Website
www.thescienceworld.net
thescienceworldmagazine@gmail.com
Published: 24.09.2023

Single molecule real time (RNAP) sequencing

This method is based on the attachment of RNA polymerase (RNAP) to a polystyrene bead, the distal end of sequenced DNA to another bead, and the placement of both beads in optical traps. The beads move closer to one another as a result of RNAP mobility during transcription, and this change in relative proximity can be captured at a single nucleotide precision. Based on the four readouts with reduced amounts of each of the four nucleotide types, the sequence is inferred.

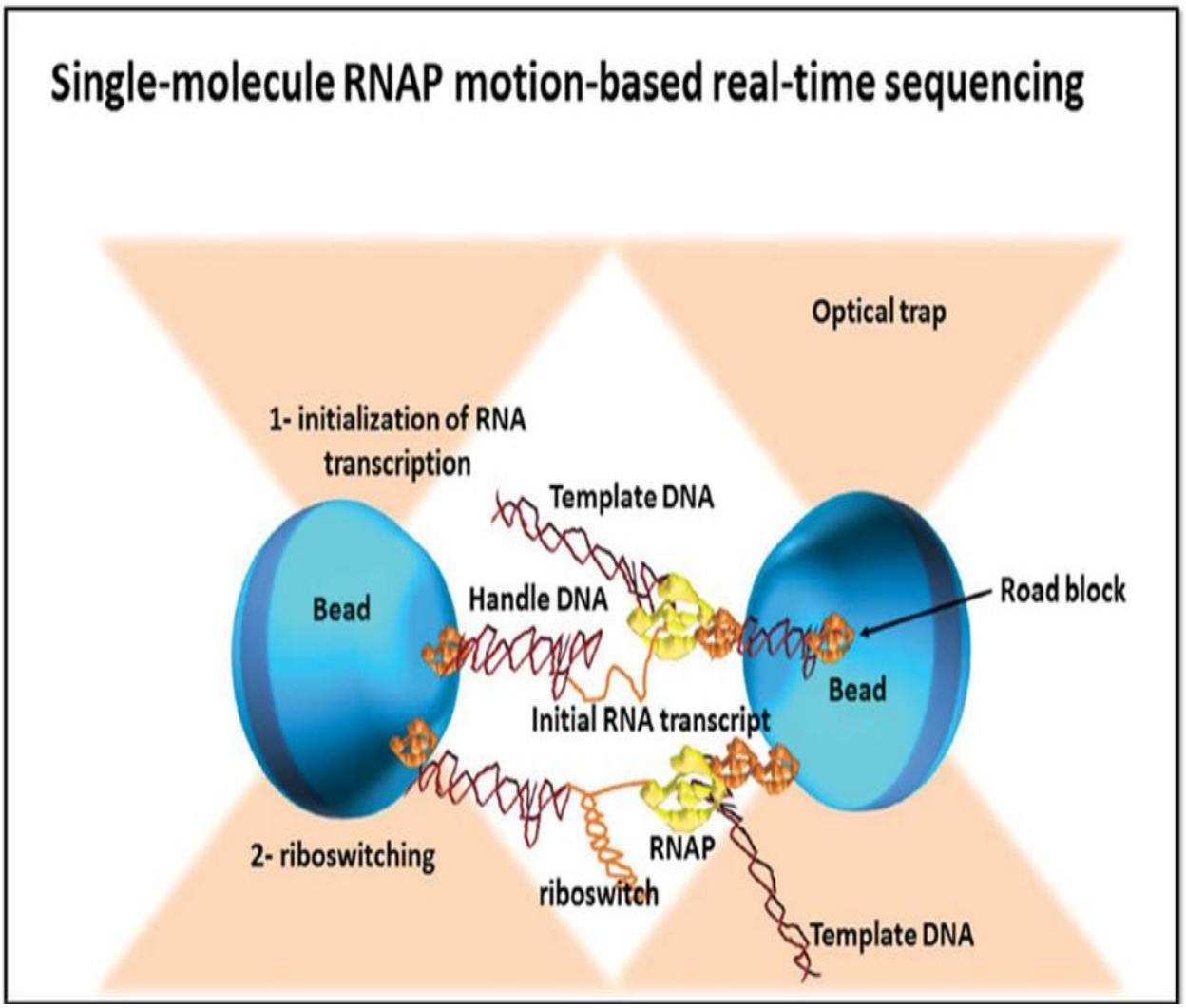


Figure 16. Single molecule real time (RNAP) sequencing

ROCHE 454 sequencing process

A) Building a system for micro-reactions. An adaptor is added to the ends of DNA fragments to create a DNA library. DNA fragments, enzymes, and dNTPs are coupled with magnetic beads and primers in the PCR reaction method. Emulsion PCR (B). By infusing water (the PCR mixture) into the oil, PCR micro reaction systems are created, and each system only comprised one template and one bead. Pyrophosphate sequencing, in (C). Pyrophosphate technology is a four-enzyme-catalyzed chemiluminescence process that takes place in the same reaction system. Only one dNTP is added to each circle of the sequencing reaction. If it only matches the following base in the DNA template, DNA polymerase will add it to the sequencing primer's 3 ′ end and release a molecular PPi at the same

Official Website
www.thescienceworld.net
thescienceworldmagazine@gmail.com
Published: 24.09.2023

time. The PPi initiates an enzyme cascade chemiluminescence reaction that is mediated by ATP sulfurylase, luciferase, and apyrase. (D) Receiving a light signal.

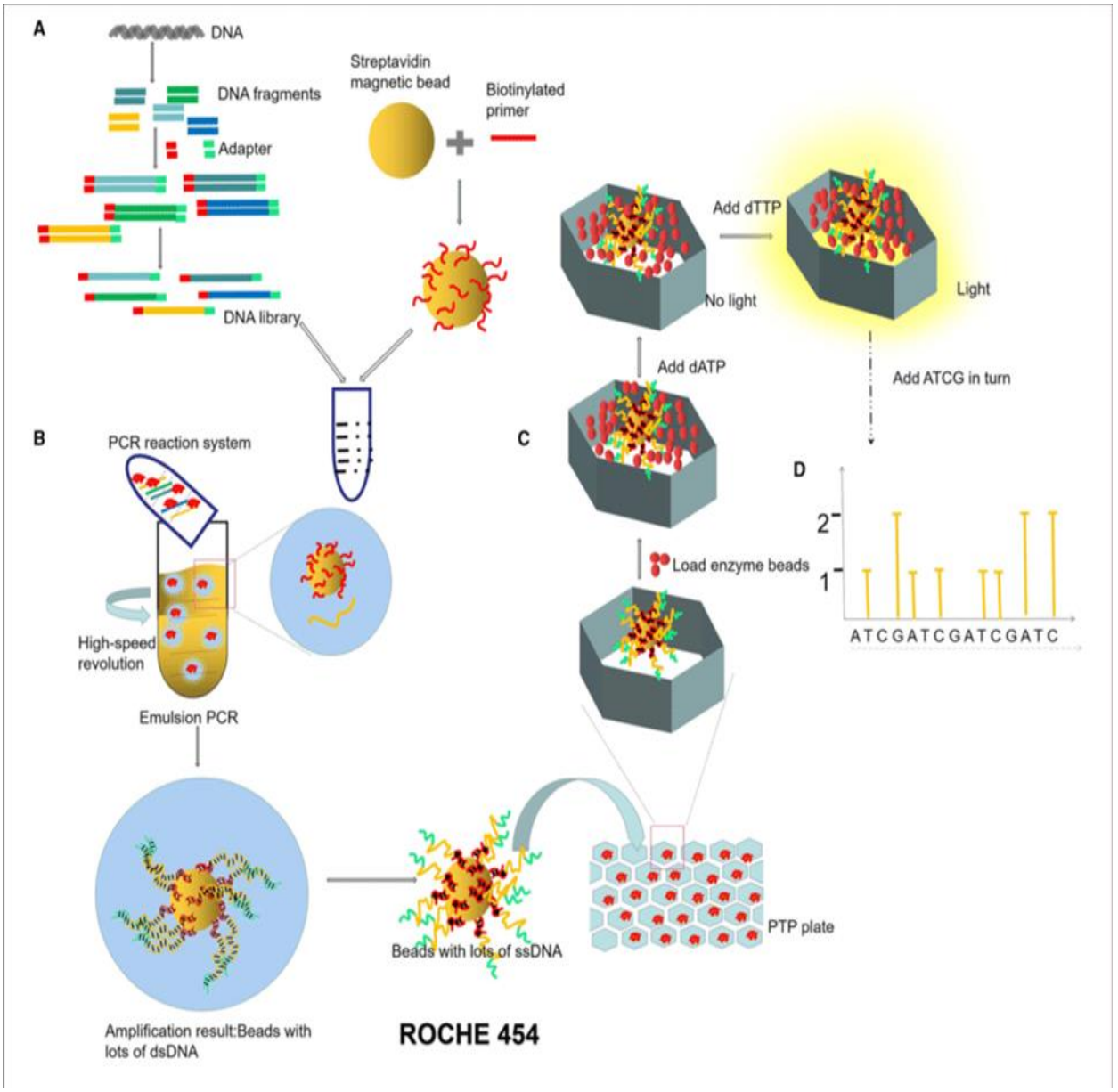


Figure 17. ROCHE 454 sequencing process

Conclusion

It is difficult to emphasize how crucial DNA sequencing is to biological study since, at its most basic, it allows us to gauge one of the key characteristics that allows terrestrial life forms to be identified and distinguished from one another. As a result, during the past 50 years, a large number of academics from all over the world have dedicated a lot of time and money to creating and advancing the technologies that support DNA sequencing. At the beginning of this field, researchers would laboriously spend years producing sequences that may range in length from a dozen to a hundred nucleotides, working mostly from accessible RNA targets. Sequencing technology has improved throughout time while also becoming more affordable thanks to advances in molecular biology, automation, and sequencing methods. This has made it possible to read DNA with hundreds of base

Official Website
www.thescienceworld.net
thescienceworldmagazine@gmail.com
Published: 24.09.2023

pairs in length and create gigabytes of data in a single run. Instead, then pouring over gels in the lab, researchers switched to running code on computers. Repositories of DNA sequence data have been expanding as genomes have been decoded, studies have been published, businesses have been founded (and frequently afterwards shut down), and so on. Therefore, DNA sequencing has a rich history despite being in many ways a modern and future-oriented academic subject.

References:

Adewale, B. A. 2020. Will long-read sequencing technologies replace short-read sequencing technologies in the next 10 years? *African journal of laboratory medicine*, *9*(1), 1-5.

Ari, Ş. and Arikan, M. 2016. Next-generation sequencing: advantages, disadvantages, and future. *Plant omics: Trends and applications*, 109-135.

Bailey, W. R., Scott, E. G., Finegold, S. M. and Baron, E. J. 1986. Bailey and Scott's Diagnostic microbiology. *(St. Louis: Mosby)*.

Cottrell, P. 2018. Advantages and Drawbacks of Next Generation Sequencing. *Available at SSRN 3183340*.

Donkor, E. S. 2013. Sequencing of bacterial genomes: principles and insights into pathogenesis and development of antibiotics. *Genes*, *4*(4), 556-572.

Hames, B. D., Hooper, N. M. and Hames, B. D. 2005. Biochemistry/David Hames and Nigel Hooper.

Deschamps, S., Llaca, V. and May, G. D. 2012. Genotyping-by-sequencing in plants. *Biology*, *1*(3), 460-483.

Head, S. R., Komori, H. K., LaMere, S. A., Whisenant, T., Van Nieuwerburgh, F., Salomon, D. R. and Ordoukhanian, P. 2014. Library construction for next-generation sequencing: overviews and challenges. *Biotechniques*, *56*(2), 61-77.

Heather, J. M. and Chain, B. 2016. The sequence of sequencers: The history of sequencing DNA. *Genomics*, *107*(1), 1-8.

http://www.aun.edu.eg/molecular_biology/Proceeding_Dec2011/DNA%20sequencing.pdf

https://ab.inf.uni-tuebingen.de/teaching/ws09/bioinformatics-i/10-sequencing.pdf

Punetha, J. and Hoffman, E. P. 2013. Short read (next-generation) sequencing: a tutorial with cardiomyopathy diagnostics as an exemplar. *Circulation: Cardiovascular Genetics*, *6*(4), 427-434.