Next-Generation DNA Sequencing: A New Era of Genetic Discovery

Kalpana Jorasia[1], Roshan Mohiddin[2], Shweta Yadav[2], Shweta Choudhary[1], Priyanka Meena[1]
[1]Post Graduate Institute of Veterinary Education and Research, Jaipur, Rajasthan, India
[2]National Dairy Research Institute, Karnal, Haryana, India
Corresponding Author: vet.shwetakhichar@gmail.com

Introduction:

DNA sequencing is the process of determining the order of nucleotides in a DNA molecule. In the study of genetics, biology, and medicine, DNA sequencing has emerged as a vital tool that has produced ground-breaking findings in areas like environmental science, agriculture, and cancer research. We will examine DNA sequencing methods in detail in this article, with an emphasis on Next Generation Sequencing (NGS), its applications, advantages, and challenges.

DNA Sequencing Techniques:

The most popular technique for sequencing DNA is the Sanger sequencing method, which was developed by Frederick Sanger in the 1970s. This technique involves the chain-termination method, where DNA synthesis is stopped by incorporating a chain- terminating nucleotide, which is detected by gel electrophoresis. However, Sanger sequencing is not feasible for large-scale sequencing projects due to its limitations, which include low throughput and high cost. These restrictions have been addressed by more recent technologies, like NGS, which are now the most popular sequencing technique.

Next Generation Sequencing (NGS):

NGS is a powerful and highly efficient DNA sequencing technology that enables the simultaneous sequencing of millions of DNA fragments. NGS involves several steps, including library preparation, sequencing, and data analysis.

Library Preparation:

The first step in NGS is library preparation, which involves breaking the DNA into smaller fragments, adding adapters to the fragments, and amplifying them to create a sequencing library. After that, the library is then loaded onto a flow cell, where the fragments are immobilized and amplified through bridge amplification.

1. **Fragmentation:** In the initial step, high-molecular-weight genomic DNA is mechanically subjected to sonication to produce smaller fragments ranging from approximately 200 to 600 base pairs (bp) in length. Subsequently, the fragmented DNA is treated with a heat-labile end-repair enzyme to generate blunt ends or phosphorylated ends.
2. **Adaptor Ligation:** In the next step, specific adaptors with complementary ends to the fragmented DNA are ligated to the ends of the DNA fragments using DNA ligase. Adaptors have several functions: they allow the fragments to bind to the sequencing platform, they provide the sequences for the primers used during the sequencing reaction, and they introduce a unique sequence for each fragment that is later used for bioinformatics analysis.
3. **Size Selection:** After adaptor ligation, the library is size-selected using gel electrophoresis or magnetic bead-based purification. The goal of size selection is to remove adaptor-dimer fragments and to select for the desired fragment size range.
4. **PCR Amplification:** The size-selected library is then amplified by PCR to generate sufficient amounts of DNA for sequencing. PCR amplification is performed using primers that anneal to the adaptor sequences and amplify the library fragments. Multiple cycles of PCR amplification are performed to generate millions of copies of the library fragments.

Sequencing:

Sequencing is performed by incorporating fluorescently labelled nucleotides into the DNA fragments, one nucleotide at a time. As each nucleotide is added, its fluorescence is detected, and the corresponding base is recorded.

1. **Clustering:** The first step in sequencing is the clustering of the library fragments on the flow cell. During clustering, each library fragment is amplified into a cluster of identical fragments on a single location on the flow cell.
2. **Sequencing-by-Synthesis:** After clustering, sequencing is performed using a sequencing-by-synthesis approach. In this method, fluorescently labelled nucleotides are incorporated into the DNA fragments one at a time, and their fluorescence is detected. Each nucleotide is identified by its unique fluorescent tag, and the sequence of the DNA fragmentis determined by the order of the nucleotides incorporated.

Illumina sequencing, Ion Torrent sequencing, Roche/454 sequencing, and SOLiD sequencing are all Next-

Generation Sequencing (NGS) technologies that have transformed genomics research.

a. **Illumina sequencing**: is also known as "sequencing by synthesis". This technology uses fluorescently-labelled nucleotides to sequence millions of DNA fragments at once. Illumina sequencing is known for its high accuracy, speed, and low cost per base.
b. **Ion Torrent sequencing,** on the other hand, is a massively parallel sequencing technology based on semi conductor technology. It detects nucleotide in corporation by measuring changes in pH as nucleotides are incorporated into a growing DNA strand. This technology is fast, easy to use, and has a low upfront cost.
c. **Roche/454 sequencing** uses pyrosequencing technology, which detects nucleotide incorporation by measuring the release of pyrophosphate as nucleotides are incorporated into a DNA strand. Roche/454 sequencing was one of the earliest NGS technologies and is known for its long reads, high accuracy, and suitability for de novo sequencing.
d. **Solid sequencing** uses reversible terminators and ligation-based chemistry to sequence DNA fragments. It is known for its high accuracy, especially in detecting small variations such as SNPs and CNVs.

Data Analysis:

Data analysis involves aligning the sequenced fragments to a reference genome, assembling the fragments.

Applications of NGS:

NGS has a wide range of applications in genomics, including whole-genome sequencing, transcriptome sequencing, and epigenome sequencing. Whole-genome sequencing is used to sequence the entire genome of an organism, while transcriptome sequencing is used to sequence the messenger RNA (mRNA) transcripts produced by an organism. Epigenome sequencing is used to determine the patterns of chemical modifications to DNA that affect gene expression.

NGS has revolutionized the field of cancer genomics by allowing for the identification of genetic mutations in tumours. This information can be used to develop personalized cancer treatments that target specific genetic mutations. NGS has also been used in agricultural research, environmental science, and microbiology.

Advantages of NGS:

NGS has several advantages over traditional sequencing methods, including high throughput, cost-effectiveness, and the ability to sequence multiple samples simultaneously. The technique also allows for the detection of rare genetic variants and the identification of novel genetic mutations.

NGS has significantly reduced the cost and time required for sequencing, making large-scale sequencing projects possible. The technology has also enabled the development of personalized medicine by allowing for the identification of genetic mutations in individual patients

Challenges of NGS:

Despite its many advantages, NGS presents several challenges. One of the main challenges is data analysis, which requires sophisticated algorithms and computational resources. The analysis of NGS data can be time-consuming and complex, requiring specialized bioinformatics expertise.

Another challenge is the quality of the data produced by NGS, which can be affected by factors such as sequencing errors, bias, and low-quality DNA samples. Quality control measures and validation techniques are required to ensure the accuracy and reliability of the results.

Conclusion:

NGS is a powerful and versatile DNA sequencing technology that has revolutionized the field of genomics. Its applications range from whole-genome sequencing to cancer genomics, and its high throughput and cost-effectiveness make it an indispensable tool for many.

* * * * * * * * *