JUST AGRICULTURE
multidisciplinary e-Newsletter

NGS Technology Used in Plant Improvement

Pranjal Patil [1] and Kartik Kharche [2]
Research Scholar, ITM University, Gwalior

ARTICLE ID: 30

Introduction

DNA sequencing is done to understand the genome's basic composition; identifying the variation inside the genome to know the variation associated with diseases gives more information regarding phenotypic traits or characters. It determines the precise order of nucleotide sequence in DNA and RNA. In recent years, crop production has shown steady and continuous growth. According to (Godfray et al. 2010) the world population will be increasing to 9 billion by the year 2050, and advanced agriculture biotechnology could be the key to improving crops to feed the world population. Therefore, in the future crop improvement productivity is necessary due to hunger and malnutrition faced by some parts/populations of the worldSequencing approaches have become increasingly popular in molecular research during the last few decades due to their technological improvement, cost-effectiveness, and high throughput. The first discovery was made by Fredrick Sanger in 1977. Sanger and his colleague introduced the 'dideoxy" chain termination method for sequencing DNA molecules. This method is also known as "Sanger Sequencing". This method was used to sequence human mitochondrial DNA – The first complete genome by Sanger and his colleagues. For over 30 years, Sanger sequencing was the most popular form of sequencing. However, Sanger sequencing, which only sequences a single DNA fragment at a time, can only sequence short pieces of DNA and is extremely time-consuming. In 1987 Leroy Hood and Michael Haunkpiller brought two major improvements to this method, firstly they labelled DNA molecules with fluorescent dye instead of radioactive molecules. Secondly, they made it possible to do sequencing work or sequencing data analysis on the computer.*NGS* is a massively parallel sequencing technology that offers high throughput, Stability, and speed to determine the order of nucleotides in the entire genome and targeted regions of RNA and DNA. The NGS plays an important role in understanding the biology of plants against the biotic and abiotic factors by applying it in genomics, Epigenomic, and transcriptomic studies. Next-generation sequencing has become an important tool for developing ideal molecular markers

JUST AGRICULTURE
multidisciplinary e-Newsletter

and identifying genes of agronomic importance. Many species including economically significant crops have undergone whole genome sequencing by de novo assembling and resequencing by utilizing next-generation sequencing methods. We now have better options to investigate crop genomics and other post-genomics (Transcriptomics, proteomics, metabolomics) in greater detail because of this NGS technology. By sequencing the entire genome of crop species and numerous accessions within crop species, we may also investigate genetic diversity and crop evaluation through the identification of molecular markers proven by NGS and other techniques. With high-throughput technology becoming more widely available and more affordable, genomics has advanced from the sequencing of small no of genes.

What is Next Generation Sequencing:

Next Generation Sequencing (NGS) is often referred to as massively parallel sequencing or high-throughput sequencing. Next-generation sequencing has drastically altered genomics and molecular biology research, allowing it to widely sequence DNA faster and at a lower cost than the Sanger method. In contrast to the Sanger method, the NGS technology sequences shorter DNA fragments, which are generally 50 to 300 nucleotides long. Although different NGS platforms (equipment used for NGS) differ substantially, the massively parallel processing of DNA fragments serves as the cornerstone for all NGS technology. Next-generation sequencers can read billions of pieces simultaneously, whereas capillary electrophoresis sequencers can only study 96 DNA or RNA fragments at a time. NGS can be used to sequence the entire genome, as well as specific reasons of interest, such as about all 20,000 coding genes or a few selected genes.

Next Generation Sequencing platform:

The NGS platform refers to the equipment used for the NGS method. Five platforms on the market differ in technology. The primary NGS platforms are Roche/454, Ion torrent, Illumina, Solid, and Single-molecule sequencing.

I. **Roche/454:** It is an important platform in NGS, it was launched in 2005. This platform uses the Pyrosequencing method which is based on the detection of pyrophosphate released during the incorporation of the new nucleotide in the newly synthesized DNA strand. The platform can generate reads with lengths of up to 1000 base pairs of

JUST AGRICULTURE
multidisciplinary e-Newsletter

genomic DNA and up to 600 base pair amplicons and can produce about 1 million reads per run.

II. **Ion torrent:** The technology used is similar to the Roche/454 platform, however instead of using pyrosequencing, it uses a hydrogen ion produced when a dNTP is incorporated, and this way the addition of new nucleotide is detected a strong positive voltage is created throughout the process due to the release of H^+ ions, which alters the pH of the surrounding the area of the picked by a device. This device can calculate 1000 base pairs of genomic DNAS and 600 base pair amplicons.

III. **Illumina:** This is the most widely used platform for next-generation sequencing it can perform extensive sequencing with high-quality readings at a lower cost. The sequencing-by-synthesis technique is used by the Illumina platform here nucleotides are recognized while also being incorporated into a nucleic acid chain during this process chemically altered nucleotides essentially bind to the DNA template strand through complementarity. These nucleotides have a florescent tag and a terminator that can be reversed to prevent the inclusion of subsequent bases. the next base can be attached to finish the sequencing of this nucleotide and the florescent signals show which new nucleotide has been added.

IV. **Solid:** This platform uses the ligation sequencing method which uses the miss-match sensitivity to determine the nucleotide sequence of a fragment. This technique uses four dyes to recognize each of the 16 potential pairings of two nucleotides with various modified probes. Two nucleotides are combined at one end of the probe, while fluorophore is located at the other end the system recognizes the light signal, and as a result the nucleotide sequence when the probe annexes the DNA sequence.

V. **Single-molecule sequencing:** This method is also known as Pacific Biosciences and the Oxford nanopore or Third-generation sequencing. The fundamental idea behind this platform is to track changes in DNA's electrical characteristics as it moves to channels the platform recognizes the pre-size nucleotide base when the fragment processes the channel. With this technique, each reaction may identify a single DNA molecule. The fundamental idea is to track changes in DNA. These modern platforms are capable of resolving issues with other sequencing techniques including reading mistakes in short reads or even the need for bench work for sample preparation.

JUST AGRICULTURE
multidisciplinary e-Newsletter

Type of DNA sequencing through NGS:

Whole Genome Sequencing, Whole Exome Sequencing and Targeted Panels are the three primary DNA sequencing techniques used by NGS.

I. **Whole Genome Sequencing (WGS):** The sequencing of an individual's entire genetic makeup or DNA is known as Whole Genome Sequencing (WGS). Both coding and non-coding DNA will be insulated in this sequencing. A very large file containing a wealth of information about the genetic material that might or might not have a clinical interpretation is produced by whole genome sequencing. However, the majority of the time only changes to the coding reason of the DNA, which makes up around 2% of the whole genome, prevent clinical interpretation. In other words, alterations that take place in non-coding regions frequently do not convey information about their potential Additionally, WGS is more expensive for processing and storing purposes to better understand the effect of changes in a place that has not yet been explored or to look for changes that lead to disease that still lack a clear couse. For example, full genome quenching.

II. **Whole Exome Sequencing (WES):** The sequencing of DNA coding region, all the DNA sequences that make up roughly 21000 genes in the human genome, is known as whole exome sequencing (WES). Exons are a group of DNA sequences that go through the process of translation transcription before being converted into proteins. Gene modification can alter protein structure which increases the chances of diseases Exome mutations are responsible for the vast majority of genetic diseases due to their effectiveness. Thus, the WES is a useful laboratory test and is better suited for clinical diagnosis, especially when there is diagnostic hypothesis of more than one particular disease or even when there is no such diagnostic hypothesis. A file containing all of the individual's gene mutations is the output of a WES which require a lengthier analysis period.

III. **Target Gene Sequencing (TGS) or Gene Panel:** Target gene sequencing, also known as a gene panel, involves doing NGS on a targeted gene. Similar to WES, these sorts of sequencing focus only on the small number of certain genes rather than the entire genome of the individual. The gene panel is therefore advised over WES in the clinical diagnosis when there is a particular disease diagnosed. This method is more beneficial than WES, as it is a cheap method of DNA sequencing, and it sequences fewer genes, so the time required for sequencing is also reduced.

JUST AGRICULTURE
multidisciplinary e-Newsletter

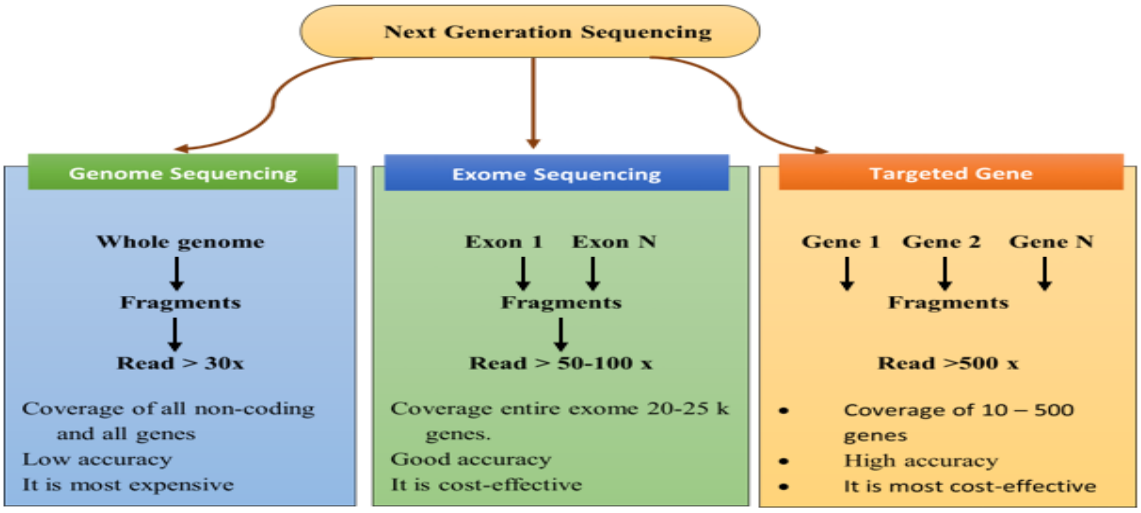

Fig 1. Types of DNA Sequencing through Next Generation Sequencing (NGS).

Steps of Next-Generation Sequencing:

There are four steps involved in Next Generation Sequencing:

- Nucleic Acid Extraction and Isolation
- Library preparation
- Clonal Amplification and Sequencing
- Data Analysis

1. **Nucleic Acid Extraction and Isolation:** Nucleic Acid Extraction and Isolation is a varying first step and essential step in next-generation sequencing. whether we are sequencing total RNA, genomic DNA (gDNA), total RNA, or other RNA forms or types. Choosing an isolating technique or kit that enables accurate cell and tissue lysis is crucial. You can then use this to get the yield, purity, and quality required for the reveling library preparation procedure.

Factors to take into account when separating nucleic acid for NGS:

- **Yield:** For library preparation nanograms (ng) to microgram(g) of DNA or RNA are typically needed therefore extracting the greatest quantity of high-quality nucleic acid

JUST AGRICULTURE
multidisciplinary e-Newsletter

is essential for the effectiveness of NGS, especially from limited sources such as cell-free DNA (cfDNA)

- **Purity:** During the library preparation isolated DNA or RNA must be free from any impurities that can inhibit the enzyme.
- **Quality:** Another crucial factor for the effectiveness of NGS is the integrity and quality of isolated nucleic acid for example; while working with gDNA, the molecular weight of separated DNA should be high and undamaged additionally isolated RNA needs to be varied, and should reflect all the nucleic acid population. Choose the proper isolation techniques to achieve enough yield and high-quality nucleic acid for sequencing.

Yield, Purity, and quality of isolated nucleic acids should be evaluated before proceeding to NGS library preparation

1. **Library Preparation:** After isolation and purification, nucleic acid is prepared for sequencer processing and reading. Since they contain a sequenceable molecule group, these ready-to-sequence samples are frequently referred to as "libraries". The steps in library preparation for the Illumine system can vary according to different techniques and the reagents used. Still, the general processes for library preparation for the Illumine system are as follows:

I. **Nucleic Acid Fragmentation:** To sequence a nucleic acid sample in a massively parallel method this sequence is broken into little pieces where the sequencer and sequencing technique determine the ideal range of fragment sizes.

II. **Adapter Ligation:** Adaptor ligation refers to the joining of oligonucleotides with complementary sequences to the sequencing chips priming oligos. To enable sequencing, adaptors are ligated onto the ends of the nucleic acid fragments. Because adaptor sequences are unique to the sequencing platform, illumine adaptors cannot be used in place of Ion torrent adaptors.

III. **Library Quantitation:** It is a pool or collection of DNA fragments with adaptors attached to their ends. To load the sequencers with the right amount of molecules to be sequenced, libraries must be prepared. This quality control phase ensures

JUST AGRICULTURE
multidisciplinary e-Newsletter

consistent data output, a high level of data quality, and effective utilization of sequencing chips. The common methods for library quantification are Fluorometric spectroscopy and real-time PCR.

DNA fragments with the necessary lengths with the adaptors of both ends makeup with the preparation of final NGS libraries.

2. **Clonal Amplification and the Sequencing reaction:**

A. Clonal amplification is further divided into two types

- Bridge Amplification.
- Cluster Generation**:**

- **Bridge Amplification:** Before sequencing, clonal amplification of DNA libraries is required to ensure that fluorescence signals from single base incorporations are detectable. Each DNA fragment from the library is coupled to primers on a sequencing chip (flow cell) via adaptors. The amplification process the Illumina platform uses solid-phase amplification, also known as bridge amplification, in which each fragment forms a cluster of identical molecules. This clonal cluster represents a single original library molecule, enabling precise signal detection. In flow cells with preset arrays, a separate approach known as Exclusion Amplification (ExAmp) is used. ExAmp swiftly amplifies a DNA fragment once it connects to the primer, preventing other DNA fragments from binding in the same spot. This approach prevents the creation of polyclonal clusters and ensures that each cluster corresponds to a single original DNA molecule.
- **Cluster Generation:** The double-stranded clonal bridges are denatured (just one strand is shown for clarity), the reverse strands are removed, and the forward strands are left as clusters for sequencing.

JUST AGRICULTURE
multidisciplinary e-Newsletter

B. Sequencing Rection: Sequencing by synthesis (SBS), which occurs after clonal amplification, is an identification of nucleotides introduced by a DNA polymerase into the complementary DNA strand of the clonal cluster. The Illumina sequencing technology uses fluorescent dye-labeled dNTPs with a reversible terminator to capture fluorescent signals in each cycle, known as cyclic reversible termination. Based on complementarity, DNA polymerase incorporates just one of the four fluorescent dNTPs every cycle, and any remaining unbound dNTPs are eliminated by washing. Following the insertion of each nucleotide, pictures of the clusters are obtained, and the base injected into each cluster during the cycle is determined by measuring the emission wavelength and fluorescence intensity of the inserted nucleotide. After imaging, the fluorescent dye and terminator are released, triggering the next cycle of synthesis, imaging, and deprotection. This method works because each base is sequenced one cycle at a time.

3. **Data Analysis:** The final step in the NGS is the processing, analysis, and interpretation of sequencing data generated. To convert raw sequencing data into meaningful results bioinformatic tools are used. The capacity and arability of computing power to process and analyze such large volumes of data are one of the workflow bottlenecks caused by NGS, which creates gig abases of raw data. After the sequencing is finished sophisticated software is utilized to interpret the vast data generated.

- **Read Processing:** Quality control processes clean and process data sequences before they are analyzed. The first stage, base calling, determines the exact nucleotide at each place in a read and assigns a quality score to each base indicating its correctness. Reads of low average quality can be deleted from the dataset. Demultiplexing then separates reads based on the adaptor and barcode sequences introduced during library preparation, sorting them according to their original sample.
- **Sequence analysis:** A comparative study of filtered sequences against a reference genome or a database comes first in the sequence analysis. Next come the tasks of removing duplicates, alignment, de novo assembly, genome annotation, variant calling, and gene prediction. Different types of bioinformatic tools can help in analysis, each having its advantages and limitations.

JUST AGRICULTURE
multidisciplinary e-Newsletter

Steps in the sequencing of data:

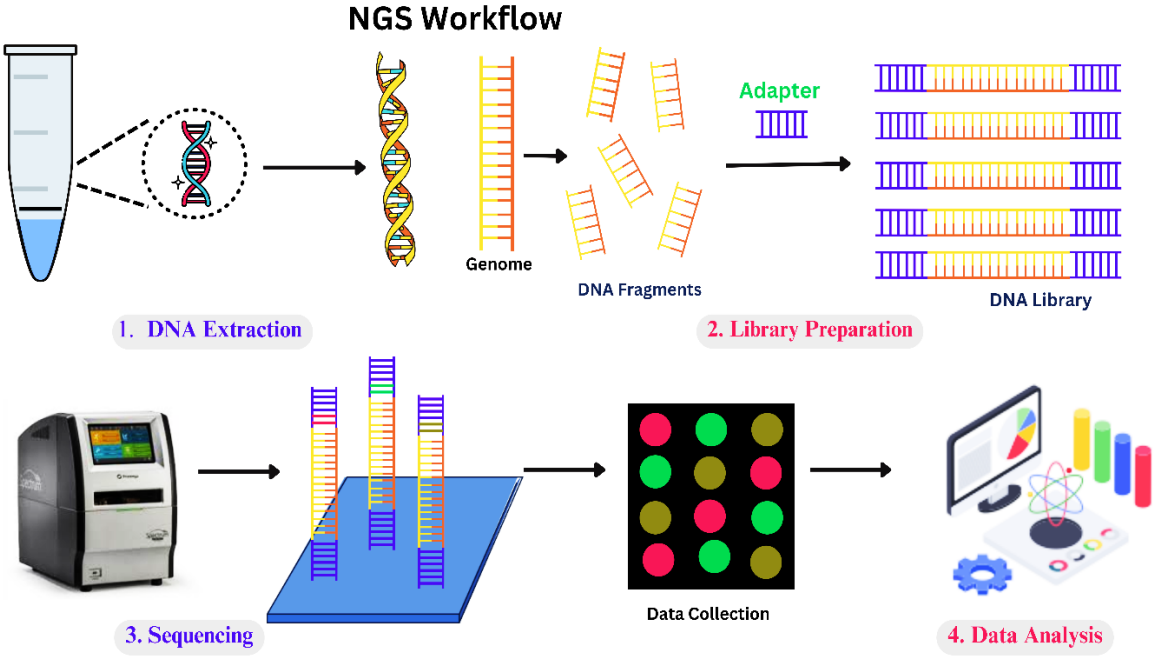

Fig 2. Diagrammatic explanation of Next-Generation sequencing (NGS) workflow.

NGS application to crop improvement:

Whole genome analysis: The rapid development of bioinformatic tools has enabled whole genome analysis on NGS sequences. Krzywinski et al. (2009) produced a Perl-written tool to detect and analyze similarities and differences between genomes using NGS sequencing alignment genome mapping and genotyping data.Breakpoint structures can be examined locally by identifying BACs that span rearrangement breakpoints and sequences that contain breakpoints. Whole genome sequencing allows you to expand and compare two genomes or chromosomes from the same species.

References

Godfray, H. C. J., Beddington, J. R., Crute, I. R., Haddad, L., Lawrence, D., Muir, J. F. & Toulmin, C. (2010). Food security: the challenge of feeding 9 billion people. *science*, *327*(5967), 812-818.

Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., ... & Marra, M. A. (2009). Circos: an information aesthetic for comparative genomics. *Genome research*, *19*(9), 1639-1645.

Kumar, K. R., Cowley, M. J., & Davis, R. L. (2024, May). Next-generation sequencing and emerging technologies. In *Seminars in thrombosis and hemostasis*. Thieme Medical Publishers.

Mandlik, J. S., Patil, A. S., & Singh, S. (2024). Next-Generation Sequencing (NGS): Platforms and Applications. *Journal of Pharmacy and Bioallied Sciences*, *16*(Suppl 1), S41-S45.

NGS workflow steps. https://sapac.illumina.com/science/technology/next-generation-sequencing/beginners/ngs-workflow.html.