THE ROLE OF NEXT-GENERATION SEQUENCING (NGS) IN BIOTECHNOLOGY: ADVANCEMENTS AND APPLICATIONS

Abstract

In the world of Biotechnology, Next-Generation Sequencing (NGS) has emerged as a revolutionary technology with profound implications for scientific research, diagnostics, and therapeutic development. This chapter explores the central role played by NGS in revolutionizing biotechnology through its remarkable advancements and diverse applications. Beginning with a discussion of the foundational principles and evolution of NGS, it highlights the technology's accessibility, working, and making high-throughput sequencing a ubiquitous tool across multiple scientific disciplines. Further, the chapter delves into the myriad applications of NGS, from unraveling the intricacies of the human genome and deciphering genetic variations, RNA sequencing, role in epigenetics, and facilitating investigations in agriculture and environmental science. It underscores NGS's transformative power in advancing scientific discovery and reshaping healthcare and beyond.

Keywords: Biotechnology, NGS, Advancement, Application.

Authors

Gourav Vats
Department of Biosciences & Technology
Maharishi Markandeshwar
(Deemed to be University)
Mullana-Ambala, Haryana, India.

Yashika Dhir
Department of Biosciences & Technology
Maharishi Markandeshwar
(Deemed to be University)
Mullana-Ambala, Haryana, India.

Sunny Dhir
Department of Biosciences & Technology
Maharishi Markandeshwar
(Deemed to be University)
Mullana-Ambala, Haryana, India.

I. INTRODUCTION

DNA sequencing methods have evolved greatly in recent years due to the advent and rapid evolution of next (NGS), also known as massively parallel sequencing [1]. NGS technologies have significantly transformed high-throughput genomics research, offering up a plethora of new fields of study and cutting-edge applications [2]. NGS technology has demonstrated tremendous promise for biologists and physicians [3]. Rapid advances in next-generation sequencing (NGS) technologies in recent years have enabled researchers to generate large numbers of sequence reads at significantly lower costs, allowing for unprecedented expansion of the scope of genome-based research projects while also revolutionizing biological and biomedical research, including studies of human disease [4-5]. Furthermore, as NGS technologies become more inexpensive, they are replacing micro array-based genotyping approaches that are confined to interrogating regions of known sequence variation [5]. This capability enabled numerous successful applications, the first of which was whole-genome sequencing (WGS), a method for sequencing the complete genome. It offers the most comprehensive landscape of genomic information and associated biological implications. Next-generation sequencing technologies are increasingly being applied in a range of disciplines. Their power stems from the ability to collect massive volumes of data and discover unique and important information about the human genome [6]. The term "NGS" refers to a group of methods that produce millions of short read sequences more quickly, more affordably, and throughput-efficiently than Sanger sequencing. The majority of NGS-based techniques that investigate genomic variation and its link to phenotypes employ a case-control study design with unrelated individuals. Because patients and controls have genetically distinct ancestries, these study designs are subject to population stratification bias (PSB) [7]. The NGS platforms that are currently available use several approaches to achieve high-throughput sequencing. The various ways of sequencing have an impact on the number, quality, and variety of sequencing applications. The extraction of genomic DNA from test samples is usually the first step in a normal NGS run. Next comes library preparation, which comprises DNA fragmentation, adaptor ligation, adaptor sequencing, sample enrichment, and sequencing. [8]. There are several NGS systems available right now [9]. Sanger sequencing, sometimes known as 'first-generation' sequencing, has been substantially displaced in recent years by technologies for 'next-generation' sequencing (NGS). It allows for the identification of biomarkers for early detection and personalized treatment. Next-generation sequencing (NGS) has altered clinical research as well as basic and applied sciences. The NGS allows for the generation of millions of data points at a lower cost. NGS has changed not only genome sequencing and personal therapy, but also previous genome research. The application of a technique that aids in establishing the amount and order of nucleotides present in the DNA of a certain organism is known as sequencing. DNA sequencing provides essential knowledge to several biological study fields as well as other industries like as biotechnology, forensic science, and biological systematic. The odd arrangement of bases has a significant impact on health, such as how we respond to a certain illness and the type of treatment necessary to cure it. NGS is a type of DNA sequencing technique that identifies the sequence by using parallel sequencing of multiple small segments of DNA [10]. High-throughput technology is quick and produces a large amount of data, far more than the Sanger technique; it is also less expensive. Third-generation sequencing is comparable to NGS in many ways, however unlike NGS; it employs single DNA molecules as a template rather than amplified DNA. As a result, third generation sequencing may eliminate mistakes in DNA sequence that occur during the DNA amplification procedure in the laboratory [11].

High-throughput technology is faster and produces significantly more data than the Sanger procedure; it is also less expensive. Third-generation sequencing is similar to next-generation sequencing in many aspects, however unlike NGS; it uses single DNA molecules as a template rather than amplified DNA. As a result, third generation sequencing may erase errors in DNA sequence that occur during the laboratory DNA amplification method [11].

II. STEPS INVOLVED IN NEXT GENERATION SEQUENCING

Step 1: Extraction of Sample: Any DNA or RNA-producing sample can be used for NGS. Depending on the starting material, there are different extraction methods that can be used. Each extraction method has been improved to yield the best quality and most nucleic acid from the relevant sample type. After extraction, evaluate the quantity and quality of DNA or RNA since successful sequencing requires high-quality starting material [12].
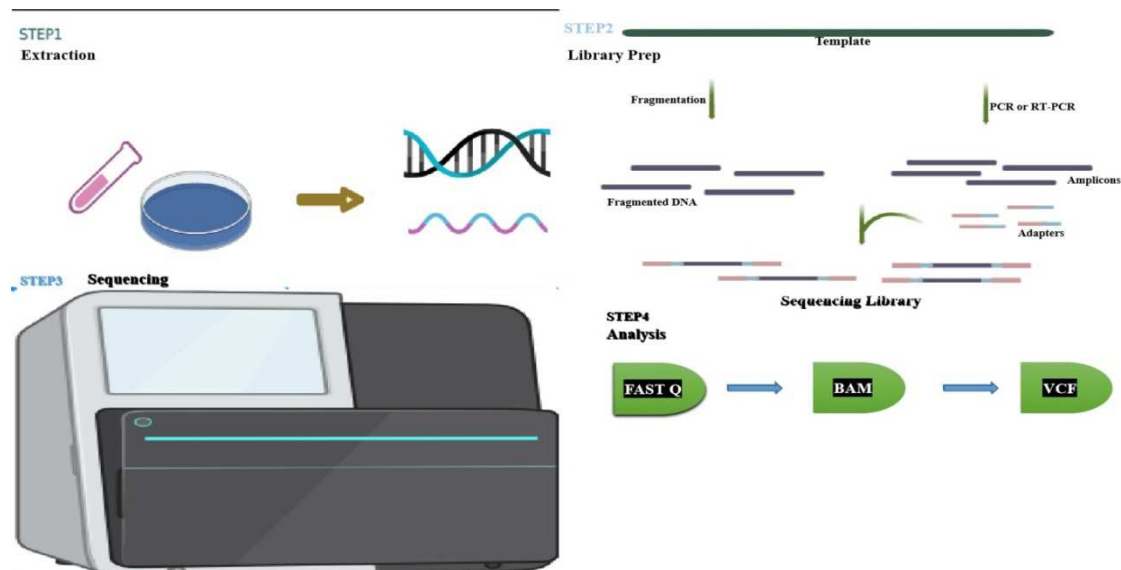
Figure 1: Figure shows various steps involved in NGS. The first step involves extraction of the genetic material to be sequenced followed by Step 2 which is library preparation. In Step 3, the library is sequenced followed by analysis in Step 4.

Step 2: Library Preparation: These are the steps required for creating a sequencing library from your RNA or DNA sample:

- Amplification to build a pool of target sequences of adequate size; and
- The incorporation of adapters for sequencing that will eventually connect to the NGS platform.

If RNA is utilized as the initial template, reverse transcription to cDNA is required. PCR amplification generates a library of DNA fragments of precise sizes that are compatible with the sequencing system to be used. Sequences of interest, which can range from the entire genome to specific RNA transcripts, are combined to create primers. Amplicons, also known as amplified DNA or cDNA fragments, are bound to particular oligonucleotide sequences at the adaptor ligation stage so that they can interface with the surface of a sequencing flow cell.

A unique identity, or barcode, is also ligated to the amplicon if a large number of samples are to be sequenced in a single run. The resulting libraries can then be combined into a single sequencing run and "demultiplexed" during the analysis of the data. [13].

Step 3: DNA Sequencing: An NGS platform is used for parallel sequencing. The sequencer "reads" the nucleotides one by one after loading the library. The amount of reads generated is determined by the sequencing platform and kit utilized. Pyrosequencing, sequencing by ligation (SOLiD), sequencing by synthesis (SBS - Illumina), and Ion Torrent sequencing are a few of the NGS methods that have been developed [14].

- **Platforms of Illumina:** Several sequencing approaches exist, despite the fact that all NGS platforms sequence millions of tiny fragments of DNA or cDNA. Some systems can generate more reads or longer reads than others. Illumina invented and popularized the most frequently used and successful sequencing technique. Illumina sequencers make use of a glass flow cell that has been coated with millions of complementary Oligonucleotides to the sequencing adaptors. Each library fragment is amplified after hybridizing with the primers, resulting in millions to billions of clonal clusters. The fluorescently labeled nucleotides are then employed to create a complimentary strand for each piece. The flow cell is photographed after each labeled nucleotide is injected, and the fluorescence from each cluster is recorded. The template sequence is determined by the wavelength and intensity of the fluorescence emission [15].

Step 4: Analysis of Data and Alignment: Following the completion of the sequencing procedure, particular software is utilized to make sense of the massive amount of data collected. Variant annotation can be used to connect variants to known genes or regulatory regions once data have been matched to a reference genome [16]. The use of NGS in healthcare, as well as variant identification and numerous gene panels will be discussed. A cohesive section discussing the essential Next Generation Sequencing characteristics in the clinic should be useful for newcomers, scientists, researchers, and health care providers who translate genetic information into genomic medicine [17].

- **Clinical Applications:** NGS has recently enabled greater understanding of genetic disorders and has emerged as a critical technological innovation in the field of detection and medical practice. NGS enables the examination of several parts of the genome in a single reaction, and it has been demonstrated to be an affordable and effective tool for investigating genetic disorders in individuals. NGS-generated genomic data has significant implications for medical practice, including the accurate identification of disease biomarkers, the detection of inherited disorders, and the finding of genetic factors that may help predict drug reactions [18]. However, clinical adoption guidelines for NGS are still being contested, restricting its application in the genetic clinic. Technology of sequencing is used in several molecular diagnostic approaches, single- and multi-gene panel tests, cell-free DNA for non-invasive prenatal testing, and whole-exome and whole-genome sequencing (WES and WGS) are examples of such assays. Due to the recent use of NGS as a diagnostic tool, questions about when to order, who to order from, and how to interpret and inform the patient and family of the results have come up. As a result, in order to identify which method is most suited to your scenario, you must first grasp its application, strength, and limitations [19].

- **Panels of Multi-Genes:** The traditional method is still highly effective for many ailments. Testing of single-gene is recommended when a patient's clinical symptoms are typical for a specific condition and the link between the disease and the individual gene is well established with little locus heterogeneity [20]. However, due to clinical diversity and genetic locus heterogeneity, many genetic disorders, X-linked intellectual disability, congenital muscular dystrophy, epilepsy, and cancer risk in families with unusual phenotypes are a few of the more challenging diagnoses, along with cardiomyopathies. [21]. The process of diagnostic has concluded, with clinical evaluation followed by sequential laboratory testing, the majority of which were negative. Undiagnosed genetic illnesses (some examples include developmental delay/cognitive disability and autism spectrum diseases) have a high diagnosis rate; hence a multi-gene panel is preferable. Tothill and colleagues show how these multi-gene panels might be used in cancer diagnoses by evaluating samples from patients with cancers of uncertain origin (CUP). The lack of a particular origin hinders clinical management of CUP patients, and this form of NGS analysis could be used to find new therapeutic possibilities. Multiple genes linked with a certain phenotype are simultaneously sequenced and assessed in multi-gene panel testing, which lowers costs and boosts the effectiveness of genetic diagnostics. [22].

III. SEQUENCING OF THE COMPLETE GENOME AND EXOME

Sequencing whole-genome is a technique for determining an organism's complete DNA sequence all at once. The primary advantage of WGS is that it covers the entire genome, including promoters and regulatory regions. Whole-exome sequencing (WES) sequences all coding regions in great detail. Coding and functional areas of the genome contain 85% of the mutations that result in illness, according to Botstein et al. Exome sequencing therefore has the potential to shed light on the aetiologies of numerous uncommon, genetically monogenic abnormalities as well as predisposing variants in prevalent illnesses and cancers [23]. In order to genetically diagnose congenital chloride diarrhea in individuals who were thought to have Bartter syndrome, a renal salt-wasting condition, Choi and colleagues employed WES in clinical practice in 2009. Six patients who lacked mutations in the classic Bartter syndrome genes underwent WES. All of the patients exhibited a homozygous deletion in the SLC26A3 gene, resulting in a molecular diagnosis of congenital chloride diarrhea that was later clinically verified Long-range PCR or Sanger sequencing. The 'known' phenotype-specific gene panels are covered more thoroughly, and compared to WES, our customized approach enables deeper coverage of these genes, boosting confidence in the discovered variations.. All NGS systems, however, are susceptible to sequencing abnormalities, and the Sanger method is advised before releasing results to patients [24]. Furthermore, all components of WES and WGS must be explained to the patient and their family. It is critical to inform them that the test might not be accurate to produce favorable results and that positive results, while providing diagnoses, do not improve prognosis or therapy.

1. **Sequencing of RNA:** The entire collection of RNA molecules from every genome at any time or place is known as a transcriptome, including untranslated RNA species like microRNAs (miRNAs), is crucial in many biological processes. RNA-sequencing (RNA-sequencing) is a transcriptomic technique that involves an in-depth RNA analysis using technologies of NGS [25]. A well-planned RNA-sequencing experiment typically includes preparation of samples, library creation, sequencing, and data interpretation.

However, due to the abundance of potential experimental possibilities, extensive preparation and expense estimation are necessary before beginning. Examples include throughput, read length, sequencing depth, and coverage. Discovering new genes and isoforms, gene fusions, splicing and chimaera variations, genomic changes, and quantifying gene expression are all made possible by RNA sequencing. [26]. Despite the fact that RNA sequencing performs better than microarray in transcriptome analysis, clinical applications are still in their infancy and won't, for instance, replace current methods. RNA-sequencing is viewed as a supplement to traditional methods, assisting doctors in making decisions based on the requirements and resources at hand. RNA analysis has practical applications in a wide range of human health domains, including drug selection, disease diagnosis, and therapy [27].

Infectious illness diagnosed clinically with RNA-sequencing is currently uncommon because the most often used approach for virus detection is quantitative PCR (RT-qPCR) and genotyping. In virology, NGS diagnostic applications can be utilized to study patients with unusual illnesses, particularly when outbreaks and epidemics occur. Additionally, it has the detection of novel pathogens, the characterization of viral communities, and the reconstruction of the full viral genome, antiviral medicine resistance, epidemiology, and transcriptome research. The use of next-generation sequencing and the application of NGS to virology are raising our understanding of the dynamics of viral infections and how they relate to human health and medicine [28, 29].

NGS is also useful for detecting circulating tumor ctRNA. The investigation of ctRNA is still present in plasma its early stages and presents distinct challenges. Because circulating tumor DNA (ctDNA) degrades faster than circulating tumor RNA (ctRNA), it must be separated rapidly or placed in preservation solutions and frozen at 80°C, a method that is not always available in many clinical settings. Despite these challenges, ctRNAs show promise for early detection of a range of tumor types, including breast, lung, prostate, and colorectal cancer. Despite the fact that NGS is a more advanced approach for detecting ctRNA, RT-qPCR is still more useful for clinical diagnostic applications [30].

2. **Epigenetics:** Epigenetics is a rapidly emerging field that has far-reaching implications for health and clinical diagnostics. In the 1940s, Conrad Waddington invented the word to describe the investigation of heritable alterations in gene expression and activity without altering the DNA sequence, i.e. a phenotype change without a genotype change. Epigenetic processes are there is yet more gene regulation, and NGS enables researchers to determine the status of epigenetics on a broad at a single base-resolution scale, particularly non-coding RNA (ncRNA)-associated silencing, methylation of DNA, and modification of histone [31]. DNA methylation was the first epigenetic process discovered, and it is the most well-known and prevalent in cancer patients. It entails methylating cytosines in CpG (cytosine/guanine) islands and covalently modifying cytosine. DNA methyltransferases (DNMTs) maintain this methylation, which is required for gene transcriptional repression, transposable element silencing, and virus protection. Active chromatin has unmethylated DNA, whereas inactive chromatin contains methylated DNA [32].

Alterations made to histones after they have been translated, for example acetylation and methylation of conserved lysines on the regions of the amino-terminal tails, are chromatin activity markers: acetylation is found in active euchromatic or heterochromatic regions, in contrast, hypo acetylation is present in inactive euchromatic or heterochromatic areas. Histone acetylases, histone deacetylases, and histone methyltransferases are some of the enzymes involved in this process. Gene expression is altered through epigenetic cell changes, which can also result in inherited diseases or acquired human illnesses because epigenetic mechanisms control DNA accessibility [33].

IV. APPLICATION OF NGS IN AGRICULTURAL BIOTECHNOLOGY FOR CROP IMPROVEMENT

1. **Genetic Modification:** The use of selection using genetic markers facilitation has considerably risen over the last thirty years. Thanks to fast genomic resource buildup, researchers now have access to an unprecedented quantity of data to access and regulate beneficial genetic variation for enhancement of crops [34]. In order to create superior cultivars for sustainable agriculture, genomic-assisted breeding is expected to increase the precision and effectiveness of breeding programmes. Ideas like genomics have emerged as a result of genotyping's incredibly high throughput and declining cost breeding-assisted genomics and breeding-assisted breeding [35]. The International Consortium of Agricultural Research Centres has adopted a new paradigm that promotes using "Omics" and interventions facilitated by bioinformatics to evaluate the level of genetic variation that is currently present, to broaden the genetic base by creating new intra- and inter-species variations, and to create novel cultivars using more efficient selection methods that combine desirable and unique features. The ultimate objective is to hasten genetic gain, which would boost food and nutritional security in developing nations while being environmentally sustainable. The enormous advances in science and technology in the fields of genetics and bioinformatics can benefit smallholder farmers in underdeveloped countries. Genetic advancement can be extremely important for increasing crop output in underdeveloped nations with limited access to agricultural inputs in a way that is environmentally sustainable. Modern genomic techniques and procedures are being used to design an efficient and productive breeding plan, driven by continually dropping genotyping costs and remarkable advancements in computational capacities. The main obstacles to breeding progress include slow genetic gain, complicated traits, and genotype by environment interaction. Before the advent of NGS, neglected crops in Africa suffered from a lack of genetic data in addition to these general limitations. Re-sequencing the entire genomes of thousands of accessions is now a practical way to access genome-wide nucleotide variation, as is using one of the complexity reduction techniques to create high-density, genome-wide SNP markers linked to important agronomic traits related to quality, environmental resistance, and biotic stresses. These technical developments allowed for the creation of experimental populations with multiple parents as well as the conventional genetic mapping inside specific biparental crossings. The efforts of IITAs (and CGIARs) to address crop productivity and other agricultural challenges have been well-documented [36].

2. **Genetic Resource Management and Utilization:** In order to protect crop genetic variety from continuous decline, gene banks are crucial. In order to continue adapting to shifting environmental conditions and consumer demands, they supply genetic variety for

breeding [37, 38]. The ability to generate enormous amounts of data from DNA sequencing technology has recently advanced, providing a chance to better explore the genetic variation preserved in the vast germplasm collections held in trust by the CGIAR and boost the effectiveness of gene banks. The 11 CGIAR gene banks preserve more than 666,000 accessions of primarily food crops [39]. The major African food crops cowpea, soybean, maize are maintained by the International Institute of Tropical Agriculture (IITA). Along with other essential crops for emerging nations, including grass pea, and their wild cousins, the aforementioned were judged understudied. To promote the use of all accessions and other genetic stocks used in breeding plans, these stocks must be thoroughly characterized [40, 41]. Although morphological descriptors have typically been used by genebanks to categorize germplasm, they are significantly influenced by the environment and different stages of plant development [42]. Additionally, there may not be as many characteristics as there should be, which makes it much harder to tell apart consanguineous kinds [43]. Technologies for molecular markers have allowed for the classification and application of a significant amount of genebank material [44]. However, due to their low prevalence in the genome, prior to the invention of NGS, marker systems that sample a portion of the genome had limited applicability. Thanks to NGS, a substantially higher density of marker analysis is now feasible. Using NGS for genotyping techniques like GBS, it has been used to investigate the genetic diversity of both farmed and wild varieties of yam. [45]. Inbreds are subjected to whole-genome fingerprinting as part of public and private breeding initiatives to better understand genetic diversity at the haplotype level [46].

3. **Breeding Data Management:** By applying cutting-edge Omics technologies, breeding projects in impoverished countries may improve the success of their breeding initiatives. There is a growing push to take advantage of improvements in the methodologies of computational biology and information and communication technology in order to effectively use a variety of morphological, genetic makeup and other datasets to supply conclusive supportive tools at various points in the breeding process. Analysis requires efficient informatics tools, automated data analysis pipelines, and decision-making tools and integration in modern breeding strategies like GS and MAS because they involve a flood of genotype data, including SNP markers procured from GBS, newly discovered statistical analysis to compute GEBV, and a lot of highly-efficient morphological data. [47]. For the efficient use of such massive quantities of genotypic, phenotypic, and other data, informatics, database, and decision support systems must be developed. Scientists in developing countries now have access to a low-cost genotyping platform thanks to multiple bilateral researches for development projects. The instruments for integrating, analyzing, and making decisions will be made easier with the help of current breeding tools and management procedures, but advancement is not possible without them. The breeding management system (BMS) of the integrated breeding platform (IBP), which was developed and marketed, is one effort that attempts to give some of these tools. The BMS service is offered by IBP regional hubs, which are hosted by collaborative research organizations like IITA in Nigeria and placed strategically throughout developing countries. The hubs promote the adoption, adaptation, and utilization of BMS and similar facilties, primarily through self-sufficiency development, technical assistance, and knowledge of particular crops. IBP already includes implemented tools and information for over 10 crops, inclusive of trait dictionaries and diagnostic markers [48, 49].

Plant breeding in the modern era and genetic study is interdisciplinary, and this multidisciplinary is supported by the collection, analysis, and use of "big data" from both field trials and laboratory investigations. Analytical chemistry is used in laboratory analysis to profile nutritional content and other metabolites, which calls for an effective data management system. An efficient data management system is required because analytical chemistry is employed in laboratory analysis to profile nutritional content and other metabolites. To make it easier to integrate various data from plant breeding, such as phenotypes from field experiments, a large open-access database containing phenotype and marker data, as well as trial design, dependable and user-friendly database is required for genotypic data, gene expression, and analytical chemistry. Additionally, a database like this must include built-in quantitative genetics analytic tools and pipelines that would enable breeders to do more than only store and retrieve raw data but also to compute breeding values and selection indices, plan crossings, and conduct field trials [50- 52].

V. ENVIRONMENTAL BAR-CODING: NEXT-GENERATION SEQUENCING

Keeping the environment healthy and the economy sustainable depend on an understanding of biodiversity that is vital to ecological study. But biodiversity science continues to be an area of unknowable. Since Linnaeus began the effort 250 years ago, more than 1.9 million species have been formally described, although it is thought that 10-100 million species currently exist on Earth [53]. As a result, not only is our depiction of biological variability laboriously lethargic, but the reality is that our best estimate of the total amount of species on Earth is subject to orders-of-magnitude uncertainty [54] shows that the tools and methodologies in use today are insufficient for making an accurate evaluation. "What species make up this ecosystem specifically?" What are the alterations in biodiversity across time, space, and in relation to upcoming environmental changes? Using biotic surveys to evaluate change in vulnerable ecosystems, biomonitoring programmes attempt to provide answers to these two fundamental issues. Both queries are challenging to respond to consistently and on time, and they are almost impossible to put into practice as monitoring targets. Because species are vulnerable to environmental disruptions that modify their natural territory, ecological authorities are increasingly employing biomonitoring approaches to assess ecosystem condition and trends. However, using conventional morphological approaches, it has proven challenging to identify taxa accurately and consistently. This is especially true for the extensive macro invertebrate sampling used in river biomonitoring, where it is frequently impossible or difficult to identify larval stages below the level of taxonomic family. Large-scale biomonitoring programmes have been difficult to execute as a result, especially in relatively less populous nations like Canada where remoteness and a lack of familiarity with the local fauna present considerable logistical challenges. The development of DNA sequencing by Sanger revolutionized every area of the biological sciences [55]. At almost any level of the taxonomic hierarchy, from individuals in populations to species to the farthest reaches of the Tree of Life, DNA sequence information can give a wealth of replicable and robust genetic data. Ecologists and evolutionary biologists now have the chance to address topics that they were unable to do so using other types of data thanks to DNA sequence-based analysis. Recent years especially after the concept of DNA barcoding was introduced in 2003 [56]. To create a common sequence library for all eukaryotes, efforts to focus DNA sequencing efforts on tiny, species-specific sections of the genome known as DNA barcodes have been successful. [57,58]. By comparing the query sequence to a DNA barcode reference library created using known species, DNA barcoding's

primary goal is to recognize unknown specimens at the species level. [59]. It is also possible to use sequence variation trends to locate novel and cryptic species. By sampling more genes or people, DNA barcode programs can shift to population-level research or in-depth phylogenetic difficulties [60]. The DNA barcode database presently has about 1.1 million individuals from over 95,000 species, all of which were contributed in the last seven decades [61]. But given that DNA barcoding has spent the last seven years primarily focused on proof-of-concept initiatives to increase application through the creation of better protocols, this advancement is substantial [62-64]. The high-throughput analysis of DNA barcodes has overcome significant obstacles, and single analytical facilities can currently analyze hundreds of thousands of samples annually. Sequence coverage in DNA barcode libraries will be significantly increased by international initiatives like the International Barcode of Life project and other coordinated efforts to barcode taxonomic groups or regional biota.

Considering that Sanger-based DNA sequencing has demonstrated its dependability for building substantial sequence libraries, including DNA barcode reference libraries, because bulk environmental samples can contain thousands of people from hundreds of species ranging from higher eukaryotes to bacteria, it is not a feasible strategy for dealing with them. Separating such individuals and utilizing single-species samples Sanger sequencing has historically been difficult and is not currently feasible using conventional procedures for some materials. Cloning and sequencing a library of cloned fragments somewhat solves this issue, but this approach has drawbacks and can induce biases. Consequently, a single-specimen Sanger sequencing technique cannot be used to routinely conduct large-scale biomonitoring programmes or other large-scale investigations of biodiversity in ecological and environmental studies. To put it another way, even while 96-well and even bigger collections of samples can be used in traditional Sanger sequencing, it is laborious to sort and segregate each individual organism into sets of 96 samples for processing. The number of organisms in a typical environmental sample ranges from hundreds to thousands, and a biomonitoring regime frequently calls for many environmental samples to be taken repeatedly over time and location. As a result, the bottleneck in this instance might not just be at the DNA sequencing phase but might also be at the steps of collecting, sorting, and preparation. It is laborious, time-consuming, and expensive to work with specimens one at a time. The 454 Genome Sequencer FLX is the primary NGS platform for biodiversity investigations because larger sequence lengths result in improved taxonomic resolution [65, 66] because it can provide sequence readings that are 250–400 bases long as opposed to the two rival platforms' less than 100 base reads. This trait is crucial since each sequencing reaction's sequenced DNA fragments (such as PCR products) would be subjected to a bioinformatics analysis in order to calculate biodiversity metrics from a set of environmental samples. Longer sequences have been found to deliver more precise biodiversity data, including species-level resolution [67]. For determining the variety of bacteria in the sample, these investigations often use sequence variation in a brief fragment of ribosomal genes. Statistical clustering techniques like BLAST are used to compare the outcomes to a relatively large sequencing library of 16S genes [68]. For large environmental samples of eukaryotic species, the same method can be used. It has been demonstrated that a short mini-barcode fragment of the mitochondrial cytochrome c oxidase 1 (COI) DNA barcode sequences—a sequence length that can be obtained easily and reliably through 454 pyrosequencing can provide the details necessary for identifying specific species with more than 90% species resolution [69-70].

VI. SUMMARY

Role of Next-Generation Sequencing (NGS) in Biotechnology: Advancements and Applications," explores how NGS technology has revolutionized biotechnology. It delves into the foundational principles and evolution of NGS, emphasizing its accessibility and cost-effectiveness. We tried to highlight the diverse applications of NGS in animal and plant biotechnology, from studying the human genome to personalized medicine, agriculture, and environmental monitoring. It also addresses methodology for data analysis, providing a comprehensive view of NGS's transformative impact on biotechnology and the life sciences.

REFERENCES

[1]   Metzker ML. Sequencing technologies – the next generation. Nat. Rev. Genet. 11(1), 31–46 (2010).

[2]   Jacquier A. The complex eukaryotic transcriptome: unexpected pervasive transcription and novel small RNAs. Nat. Rev. Genet. 10(12), 833–844 (2009).

[3]   Shendure J, Ji H. Next-generation DNA sequencing. Nat. Biotechnol. 26(10), 1135–1145 (2008).

[4]   Ng SB, Turner EH, Robertson PD, et al. Targeted capture and massively parallel sequencing of 12 human exomes. Nature. 2009;461:272–6.

[5]   Ng SB, Buckingham KJ, Lee C, et al. Exome sequencing identifies the cause of a Mendelian disorder. Nat Genet. 2010;42:30–5.

[6]   Cirulli ET, Goldstein DB. Uncovering the roles of rare variants in common disease through whole genome sequencing. Nat Rev Genet. 2010;11:415–25.

[7]   Freedman, M. L., Reich, D., Penney, K. L., McDonald, G. J., Mignault, A. A., Patterson, N., et al. (2004). Assessing the impact of population stratification on genetic association studies. Nat. Genet. 36, 388-393. doi: 10.1038/ng1333

[8]   Buermans, H. P. J., and den Dunnen, TJ. (2014). Next generation sequencing technology: advances and applications. Biochim. Biophys. Acta 1842, 1932-1941. doi: 10.1016/j.bbadis.2014.06.015

[9]   Heather, J. M., and Chain, B. (2016). The sequence of sequencers: the history of sequencing DNA. Genomics 107, 1-8. doi: 10.1016/j.ygeno.2015.11.003

[10]   Rizzo, J.M. and Buck, M.J., 2012. Key principles and clinical applications of "next- generation" DNA sequencing. Cancer prevention research, 5(7), pp.887-900.

[11]   Nakada, S., Tai, I., Panier, S., Al-Hakim, A., Iemura, S.I., Juang, Y.C., O'Donnell, L., Kumakubo, A., Munro, M., Sicheri, F. and Gingras, A.C., 2010. Non-canonical inhibition of DNA damage-dependent ubiquitination by OTUB1. Nature, 466(7309), pp.941-946.

[12]   Hart, M.L., Meyer, A., Johnson, P.J. and Ericsson, A.C., 2015. Comparative evaluation of DNA extraction methods from feces of multiple host species for downstream next- generation sequencing. PloS one, 10(11), p.e0143334.

[13]   Van Dijk, E.L., Jaszczyszyn, Y. and Thermes, C., 2014. Library preparation methods for next-generation sequencing: tone down the bias. Experimental cell research, 322(1), pp.12- 20.

[14]   Foox, J., Tighe, S.W., Nicolet, C.M., Zook, J.M., Byrska-Bishop, M., Clarke, W.E., Khayat, M.M., Mahmoud, M., Laaguiby, P.K., Herbert, Z.T. and Warner, D., 2021. Performance assessment of DNA sequencing platforms in the ABRF Next-Generation Sequencing Study. Nature biotechnology, 39(9), pp.1129-1140.

[15]   Hendre, P.S., Kamalakannan, R. and Varghese, M., 2012. High throughput and parallel SNP discovery in selected candidate genes in Eucalyptus camaldulensis using Illumina NGS platform. Plant biotechnology journal, 10(6), pp.646-656.

[16]   Hendre, P.S., Kamalakannan, R. and Varghese, M., 2012. High throughput and parallel SNP discovery in selected candidate genes in Eucalyptus camaldulensis using Illumina NGS platform. Plant biotechnology journal, 10(6), pp.646-656.

[17]   Schadt, E.E., Turner, S. and Kasarskis, A., 2010. A window into third-generation sequencing. Human molecular genetics, 19(R2), pp.R227-R240.

[18]   Jamuar, S.S. and Tan, E.C., 2015. Clinical application of next-generation sequencing for Mendelian diseases. Human genomics, 9(1), pp.1-6.

[19]   Rabbani, A., Zamani, M., Yazdani-Chamzini, A. and Zavadskas, E.K., 2014. Proposing a new integrated model based on sustainability balanced scorecard (SBSC) and MCDM approaches by using linguistic

variables for the performance evaluation of oil producing companies. Expert systems with applications, 41(16), pp.7316-7327.

[20] Xue, Y., Ankala, A., Wilcox, W.R. and Hegde, M.R., 2015. Solving the molecular diagnostic testing conundrum for Mendelian disorders in the era of next-generation sequencing: single-gene, gene panel, or exome/genome sequencing. Genetics in Medicine, 17(6), pp.444-451.

[21] LaDuca, H., Farwell, K.D., Vuong, H., Lu, H.M., Mu, W., Shahmirzadi, L., Tang, S., Chen, J., Bhide, S. and Chao, E.C., 2017. Exome sequencing covers> 98% of mutations identified on targeted next generation sequencing panels. Plos one, 12(2), p.e0170843.

[22] Rossello, F.J., Tothill, R.W., Britt, K., Marini, K.D., Falzon, J., Thomas, D.M., Peacock, C.D., Marchionni, L., Li, J., Bennett, S. and Tantoso, E., 2013. Next-generation sequence analysis of cancer xenograft models. PloS one, 8(9), p.e74432.

[23] Song, W., Zhu, D., Lv, Y. and Wang, W., 2016. Development and characterization of 29 polymorphic microsatellite loci of Megalobrama pellegrini by next-generation sequencing technology and cross-species amplification in related species (No. e2490v1). PeerJ Preprints.

[24] Salgotra, R.K., Gupta, B.B. and Stewart, C.N., 2014. From genomics to functional markers in the era of next-generation sequencing. Biotechnology letters, 36, pp.417-426.

[25] Sen, S.K., Barb, J.J., Cherukuri, P.F., Accame, D.S., Elkahloun, A.G., Singh, L.N., Lee-Lin, S.Q., Program, N.C.S., Kolodgie, F.D., Cheng, Q. and Zhao, X., 2014. Identification of candidate genes involved in coronary artery calcification by transcriptome sequencing of cell lines. BMC genomics, 15, pp.1-12.

[26] Zhang, W., Yu, Y., Hertwig, F., Thierry-Mieg, J., Zhang, W., Thierry-Mieg, D., Wang, J., Furlanello, C., Devanarayan, V., Cheng, J. and Deng, Y., 2015. Comparison of RNA-seq and microarray-based models for clinical endpoint prediction. Genome biology, 16(1), pp.1- 12.

[27] Byron, S.A., Van Keuren-Jensen, K.R., Engelthaler, D.M., Carpten, J.D. and Craig, D.W., 2016. Translating RNA sequencing into clinical diagnostics: opportunities and challenges. Nature Reviews Genetics, 17(5), pp.257-271.

[28] Sijmons, S., Van Ranst, M. and Maes, P., 2014. Genomic and functional characteristics of human cytomegalovirus revealed by next-generation sequencing. Viruses, 6(3), pp.1049- 1072.

[29] Xiong, Y., Chen, X., Chen, Z., Wang, X., Shi, S., Wang, X., Zhang, J. and He, X., 2010. RNA sequencing shows no dosage compensation of the active X-chromosome. Nature genetics, 42(12), pp.1043-1047.

[30] Bray, N., Pimentel, H., Melsted, P. and Pachter, L., 2015. Near-optimal RNA-Seq quantification. arXiv preprint arXiv:1505.02710.

[31] Egger, G., Liang, G., Aparicio, A. and Jones, P.A., 2004. Epigenetics in human disease and prospects for epigenetic therapy. Nature, 429(6990), pp.457-463.

[32] Skinner, M.K., 2011. Role of epigenetics in developmental biology and transgenerational inheritance. Birth Defects Research Part C: Embryo Today: Reviews, 93(1), pp.51-55.

[33] Baerwald, M.R., Meek, M.H., Stephens, M.R., Nagarajan, R.P., Goodbla, A.M., Tomalty, K.M., Thorgaard, G.H., May, B. and Nichols, K.M., 2016. Migration related phenotypic divergence is associated with epigenetic modifications in rainbow trout. Molecular Ecology, 25(8), pp.1785-1800.

[34] Heslot N, Sorrels ME, Jannink J. Perspectives for genomic selection applications and research in plants. Crop Sci 2015;55:1–30. doi:10.2135/cropsci2014.03.0249.

[35] Poland J. Breeding-assisted genomics. Curr Opin Plant Biol 2015;24:119–24. doi: 10.1016/j.pbi.2015.02.009.

[36] Gedil M, Tripathi L, Ghislain M, Ferguson M, Ndjiondjop M, Kumar PL, et al. Biotechnology success stories by the Consultative Group on International Agriculture Research (CGIAR) system. In: Wambugu F, Kamanga D, editors. Biotechnology in Africa: Emergence, Initiative and Future. Springer, Heidelberg. 2014, p. 95-114.

[37] Bansal KC, Lenka SK, Mondal TK. Genomic resources for breeding crops with enhanced abiotic stress tolerance. Plant Breed 2014;133:1–11. doi:10.1111/pbr.12117.

[38] Dwivedi SL, Upadhyaya HD, Stalker HT, Blair MW, Bertioli DJ, Nielen S, et al. Enhancing crop gene pools with beneficial traits using wild relatives. Plant Breed Rev 2008;30:179–230.

[39] Koo B, Pardey PG, Wright BD. The economic costs of conserving genetic resources at the CGIAR centers. Agric Econ 2003;29:287–97. doi:10.1111/j. 1574-0862.2003.tb00165.x.

[40] Lee S-H, Tuberosa R, Jackson SA, Varshney RK. Genomics of plant genetic resources: a gateway to a new era of global food security. Plant Genet Resour 2014;12:S2–5. doi: 10.1017/S1479262114000513.

[41] Romay MC, Millard MJ, Glaubitz JC, Peiffer JA, Swarts KL, Casstevens TM, et al. Comprehensive genotyping of the USA national maize inbred seed bank. Genome Biol 2013;14:R55. doi:10.1186/gb-2013-14-6-r55.

[42] Elhoumaizi MA, Saaidi M, Oihabi A, Cilas C. Phenotypic diversity of date-palm cultivars(*Phoenix dactylifera* L.) from Morocco. Genet Resour Crop Evol 2002;49:483–90. doi:10.1023/A:1020968513494.

[43] Duminil J, Di Michele M. Plant species delimitation: a comparison of morphological and molecular markers. Plant Biosyst – An Int J Deal with All Asp Plant Biol 2009;143:528–42. doi:10.1080/11263500902722964.

[44] Börner A, Khlestkina EK, Chebotar S, Nagel M, Arif MAR, Neumann K, et al. Molecular markers in management of ex situ PGR – A case study. J Biosci 2012;37:871–77. doi:10.1007/s12038-012-9250-2.

[45] Girma G, Hyma KE, Asiedu R, Mitchell SE, Gedil M, Spillane C. Next-generation sequencing based genotyping, cytometry and phenotyping for understanding diversity and evolution of guinea yams. Theor Appl Genet 2014;127:1783–94.

[46] Padi FK, Ofori A, Takrama J, Djan E, Opoku SY, Dadzie AM, et al. The impact of SNP fingerprinting and parentage analysis on the effectiveness of variety recommendations in cacao. Tree Genet Genomes 2015;11:44. doi:10.1007/s11295-015-0875-9.

[47] Wallace JG, Upadhyaya HD, Vetriventhan M, Buckler ES, Tom Hash C, Ramu P. The genetic makeup of a Global Barnyard Millet Germplasm Collection. Plant Genome 2015;8:1-7. doi:10.3835/plantgenome2014.10.0067.

[48] McCouch S, Baute GJ, Bradeen J, Bramel P, Bretting PK, Buckler E, et al. Agriculture: feeding the future. Nature 2013;499:23–24. doi:10.1038/499023a.

[49] Girma G, Korie S, Dumet D, Franco J. Improvement of accession distinctiveness as an added value to the global worth of the yam (*Dioscorea* spp) genebank. Int J Conserv Sci 2012;3:199–206.

[50] Galperin MY, Rigden DJ, Fernández-Suárez XM. The 2015 Nucleic Acids Research Database Issue and molecular biology database collection. Nucleic Acids Res 2015;43:D1–5. doi:10.1093/nar/gku1241.

[51] Shirasawa K, Isobe S, Tabata S, Hirakawa H. Kazusa Marker DataBase: a database for genomics, genetics, and molecular breeding in plants. Breed Sci 2014;64:264–71. doi: 10.1270/jsbbs.64.264.

[52] Grant D, Nelson RT, Cannon SB, Shoemaker RC. SoyBase, the USDA-ARS soybean genetics and genomics database. Nucleic Acids Res 2010;38: D843–46. doi:10.1093/nar/ gkp798.

[53] Schaeffer ML, Harper LC, Gardiner JM, Andorf CM, Campbell DA, Cannon EKS, et al. MaizeGDB: curation and outreach go hand-in-hand. Database (Oxford) 2011;2011:bar022. doi:10.1093/database/bar022.

[54] Chapman, A.D., 2009. Numbers of living species in Australia and the world.

[55] Wilson EO (1985) Time to Revive Systematics. Science 230: 1227.

[56] Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci U S A 74: 5463–5467.

[57] Hebert PDN, Cywinska A, Ball SL, deWaard JR (2003) Biological identifications through DNA barcodes. Proc R Soc Lond B Biol Sci 270: 313–321.

[58] Marshall E (2005) Taxonomy. Will DNA bar codes breathe life into classification? Science 307: 1037.

[59] Savolainen V, Cowan RS, Vogler AP, Roderick GK, Lane R (2005) Towards writing the encyclopaedia of life: an introduction to DNA barcoding. Philos Trans R Soc Lond B Biol Sci 360: 1805–1811.

[60] Hajibabaei M, Singer GAC, Hebert PDN, Hickey DA (2007) DNA barcoding: how it complements taxonomy, molecular phylogenetics and population genetics. Trends Genet 23: 167–172.

[61] Ratnasingham S, Hebert PDN (2007) BOLD: The Barcode of Life Data System (www.barcodinglife.org). Mol Ecol Notes 7: 355–364.

[62] deWaard JR, Ivanova NV, Hajibabaei M, Hebert PDN (2008) Assembling DNA barcodes. Analytical protocols. Methods Mol Biol 410: 275–293.

[63] Hajibabaei M, dewaard JR, Ivanova NV, Ratnasingham S, Dooh RT, et al. (2005) Critical factors for assembling a high volume of DNA barcodes. Philos Trans R Soc Lond B Biol Sci 360: 1959–1967.

[64] Hajibabaei M, Smith MA, Janzen DH, Rodriguez JJ, Whitfield JB, et al. (2006) A minimalist barcode can identify a specimen whose DNA is degraded. Mol Ecol Notes 6: 959–964.

[65] Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, et al. (2005) Genome sequencing in microfabricated high-density picolitre reactors. Nature 437: 376–380.

[66] Sogin ML, Morrison HG, Huber JA, Welch DM, Huse SM, et al. (2006) Microbial diversity in the deep sea and the underexplored "rare biosphere". Proc Natl Acad Sci U S A 103: 12115–12120.

[67] Meusnier I, Singer GA, Landry JF, Hickey DA, Hebert PDN, et al. (2008) A universal DNA mini-barcode for biodiversity analysis. BMC Genomics 9: 214.

[68] Ley RE, Turnbaugh PJ, Klein S, Gordon JI (2006) Microbial ecology: human gut microbes associated with obesity. Nature 444: 1022–1023.

[69] Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215: 403–410.

[70] Hajibabaei M, Singer GA, Clare EL, Hebert PDN (2007) Design and applicability of DNA arrays and DNA barcodes in biodiversity monitoring. BMC Biol 5: 24.