JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Comprehensive review on Next Generation Sequencing methodology and its applications

Deepthi V*, Ramya G.N.V.S, Nandini P and Reddy O.V.S.
Department of Biochemistry, Sri Venkateswara University, Tirupati – 517 502, India

Abstract: The main output of the next generation sequencing is determination of sequence data of DNA or RNA in a vast amount. This sequencing is represented as a series of bases of nucleotides-Adenine, Thymine, Cytosine, Guanine for DNA and Adenine, Uracil, Guanine, Cytosine for RNA. This provides the genetic code of sample which is being sequenced. The first-generation sequencing such as Sanger sequencing analysis which is a time-consuming method when compared with second generation sequencing known as NGS. It may take time from few hours to some days but NGS complete the data sequence of DNA or RNA with in a single day (1). NGS also known as High-Throughput Sequencing (HTS), Massively Parallel Sequencing (MPS), Second generation sequencing etc. This review reflects the importance of NGS methodology and its applications in a simple manner. NGS, a revolutionary advancement over the first-generation Sanger sequencing, enables rapid and comprehensive sequencing of DNA and RNA. NGS deciphers the genetic code of samples with unprecedented speed and efficiency. This review underscores the significance of NGS technology and explores its wide-ranging applications, highlighting its transformative impact on genomic research and diagnostics.

Keywords: NGS methodology, Applications, Advantages and Disadvantages.

Introduction

In the early 21st century the ground breaking technology was proposed which plays a significant role in DNA sequencing known as Next Generation Sequencing (NGS).The introduction of first DNA sequencing method took longer time (25 years) after the discovery of DNA structure by Watson and Crick in the year of 1953 (2, 3). There is a difference between NGS method and Sanger sequencing method. In case of NGS: provides massively parallel analysis, high-throughput from multiple samples at reduced cost (4). Whereas Sanger sequencing method is used to sequence both tiny as well as larger size genomes from bacteria to humans (5).

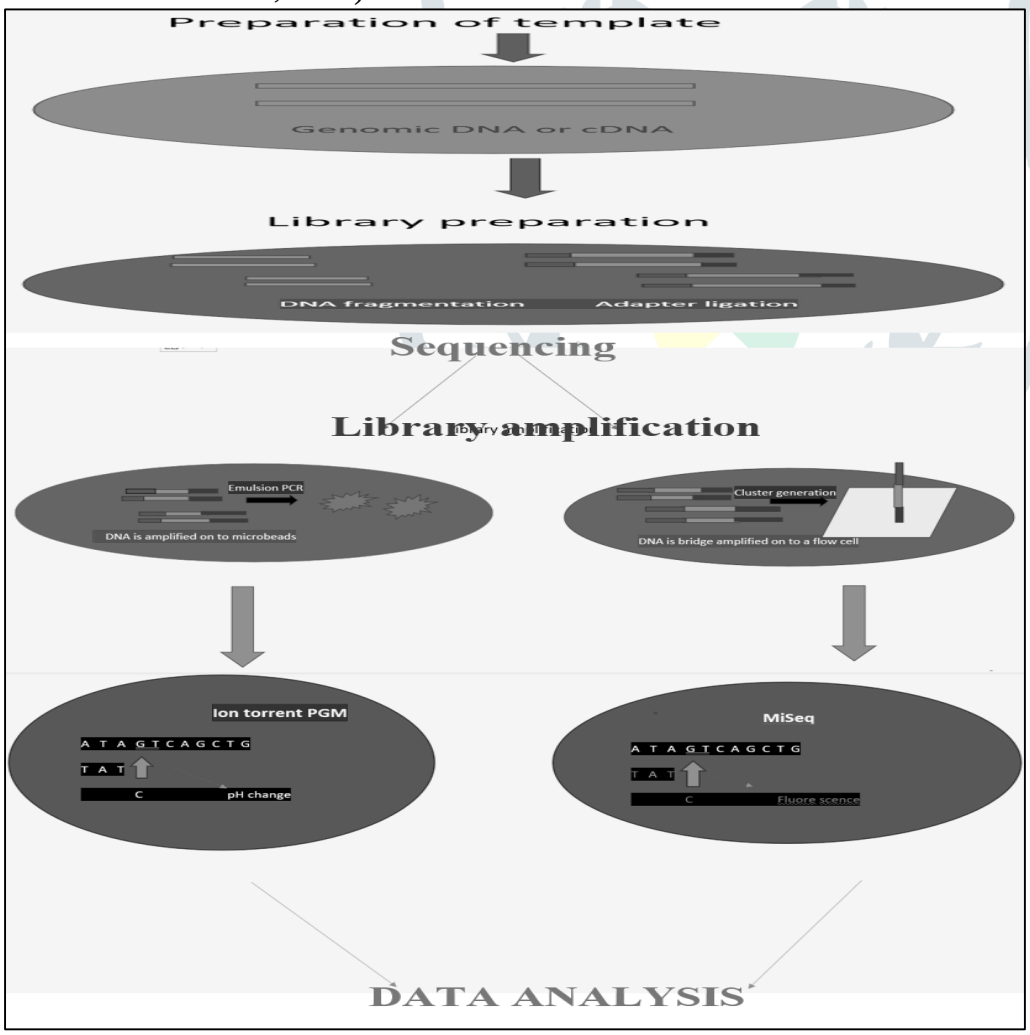
There are many applications of NGS in the development of Biotechnology, Bioinformatics and molecular biology. Also used in clinical laboratories including DNA testing, typing of Human Leucocyte Antigen (HLA), for the analysis of microbes and gene expression. For the testing of identity by using of short tandem repeats (STRs). At current the most clinically available earliest development is NGS of circulating tumor DNA (ctDNA). Advanced blood group typing like - TypeA1 versus Type A2 (6). This review mainly focuses on methodology of next generation sequencing and its applications and briefly reflects about advantages, and disadvantages. The impact of NGS extends to clinical settings, where it is increasingly used for diagnostic purposes, personalized medicine, and cancer genomics. The ability to rapidly and accurately identify genetic mutations associated with diseases has opened new avenues for treatment and management, highlighting the critical role of NGS in modern healthcare. Despite its remarkable advancements, NGS faces challenges such as data management complexities, the need for high computational power and interpretation of vast datasets. Ongoing improvements in sequencing technologies, bioinformatics tools and data analysis methods are essential to fully harness the potential of NGS. It has emerged as a groundbreaking technology that has revolutionized the field of genomics by providing rapid, accurate, and comprehensive sequencing capabilities. Unlike traditional Sanger sequencing, which sequences DNA one fragment at a time, NGS can sequence millions of fragments simultaneously. This parallel processing enables researchers to achieve a high throughput at significantly lower costs and shorter turnaround. NGS has profoundly impacted biological research and clinical diagnostics, providing deep insights into genetic and molecular mechanisms. As technologies continue to evolve, NGS is expected to become even more integral to personalized medicine, enabling more precise and effective healthcare solutions.

Methodology of NGS: There are two main platforms of next generation sequencing techniques which are commonly used for DNA sequencing. They are PGM (Ion Personal Genome Machine) and Illumina MiSeq. These are related to broad category of technique which revolutionized sequencing of DNA compared to traditional Sanger sequencing. NGS provides high-throughput of DNA sequencing, enabling simultaneous analysis of abundant sequencing are generated in parallel, significantly rises the velocity and proficiency of sequencing process. Even though the main aim of these two plat-forms are same they are differing in their tools using in sequencing techniques and are shown in the Table-1.

Table. 1 Disparity between PGS and MiSeq (Ayman Grada et al., 2001).

PGM	MiSeq
1. Stands for personal genome machine	1. Stands for microbial sequencing
2. Developed by life technologies	2. Developed by Illumina
3. Detect the changes in pH as nucleotides are incorporated during synthesis of DNA by using semiconductor sequencing technology	3. Detect the sequencing by fluorescent labelled nucleotides are added to the growing DNA strand and detected by using reversible terminated base technology

There are various steps involved in next generation sequencing. They are 1. Preparation of template 2. Library preparation 3. Library amplification 4. Sequencing and Data analysis shown in Figure. 1 (Ayman Grada and Kate Weinbrecht, 2013).



Preparation of template and library preparation: Template preparation creates a fragment of DNA or RNA for analysis. (a) Extraction of nucleic acid of interest. (B) Fragmentation of extracted DNA into miniature pieces by using sonication (or) other enzymatic technique. (C) Small pieces of DNA fragments are repaired which provide blunt ends. In this step the adopters can be simply ligated to the ends. (d) The bases of adenosine (A) are

added to 3' end of the fragments of DNA which is known as A-Tailing. And follows a size exclusion method to obtain a uniform library.

Library amplification: This step is used to create clumps of similar DNA fragments. Here PGM and MiSeq use different tools for amplification. For the amplification of library fragments on to microbeads PGM use emulsion PCR on the one touch system and the bridge amplification technique is used by MiSeq for the formation of clusters of samples on a flow cell. (Berglund et al, 2011: Quail et al, 2012). The Ion Torrent PGM and Illumina's MiSeq are two popular NGS platforms that utilize different techniques for library amplification.

Ion Torrent PGM: This system uses emulsion PCR on the OneTouch system to amplify library fragments onto microbeads. In this technique, DNA fragments are captured on microbeads within tiny water droplets in an oil emulsion. Each droplet acts as an individual micro-reactor where PCR amplification occurs, resulting in thousands of copies of the original DNA fragment bound to a single bead. This process ensures a high degree of uniformity and density, which is crucial for the subsequent sequencing steps. The emulsion PCR method is particularly advantageous for its ability to handle a high throughput of samples and produce robust results, as detailed by Berglund et al. (2011).

Illumina MiSeq: In contrast, the MiSeq platform employs bridge amplification to form clusters of DNA fragments on a flow cell. This method involves the immobilization of DNA fragments onto a solid surface where they undergo amplification. The DNA fragments first bind to complementary adapter sequences that are fixed on the flow cell surface. Through a series of amplification cycles, these fragments bend and form bridges, allowing for the synthesis of new strands that are tethered to the flow cell. This results in dense clusters of identical DNA fragments, enhancing the signal strength during sequencing. Bridge amplification is known for its efficiency and high accuracy, making it a preferred technique for generating high-quality sequencing data, as noted by Quail et al. (2012).

Both emulsion PCR and bridge amplification are essential for ensuring that the sequencing process is accurate and reliable, allowing for the detailed analysis of complex genomes. The choice between these methods typically depends on the specific requirements of the sequencing project, such as the type of samples, desired throughput, and the platform being used. These amplification techniques play a critical role in the overall success of next-generation sequencing, contributing significantly to advancements in genomic research and personalized medicine.

Sequencing: The new fragments of DNA are assembled by using template which are library fragments. In the growing DNA strand new nucleotides are incorporate which are digitally observed as nucleotide sequence. During the incorporation of new nucleotides in the growing strand of DNA the changes in values of pH takes place. These pH changes are identified by using semiconductor sequencing platform is used in case of PGM, whereas the MiSeq method which is complimentary to the PGM. It detects fluorescence of fluorescently labelled nucleotides by using reversible terminated base technology.

Data analysis: This is the final step of next generation method which involves the processing and interpreting large quantity of raw sequence data generated during the process of sequencing. There are various steps are used for the analysis like base calling, Quality control, read alignment, variant calling, assembly (if applicable), annotation, functional analysis, statistical analysis, data visualization (7-16).

Base calling: This initial step converts raw signal data from the sequencing instrument into nucleotide sequences, identifying the sequence of bases (A, T, C, and G) in each DNA fragment.

Quality control: Ensuring the integrity and reliability of sequencing data, quality control involves assessing the quality of the sequences and filtering out low-quality reads. This step often employs tools like FastQC to visualize and evaluate the data.

Read alignment: The high-quality sequences are then aligned to a reference genome. This process, also known as mapping, determines where each read fits in the context of the entire genome, using algorithms and software such as BWA or Bowtie.

Variant calling: This step identifies genetic variations, such as single nucleotide polymorphisms (SNPs) and insertions/deletions (indels), by comparing the aligned sequences to the reference genome. Tools like GATK or Samtools are commonly used for this purpose.

Assembly (if applicable): In the absence of a reference genome, *de novo* assembly is performed to reconstruct the genome sequence from short reads. Software like Spades' or Velvet can assemble these sequences into longer contiguous sequences, or contigs.

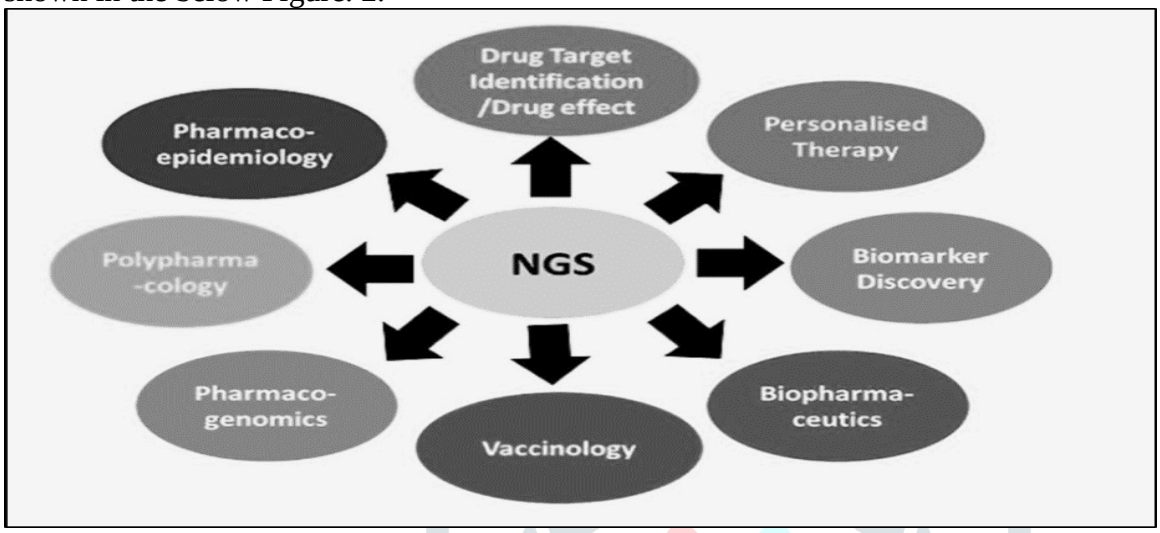
Annotation: Annotation involves identifying and marking features within the genome, such as genes, regulatory elements and other functional regions. This process uses databases like Ensemble or NCBI for reference.

Functional analysis: This step interprets the biological significance of the annotated features, linking genetic variations to potential functional impacts. It may include pathway analysis and gene ontology enrichment to understand the broader biological implications.

Statistical analysis: Rigorous statistical methods are applied to validate the findings and assess their significance. This ensures that the observed variations and patterns are not due to random chance, but rather represent true biological phenomena.

Data visualization: Finally, data visualization techniques present the analysis results in an accessible and interpretable format. Visualization tools like IGV (Integrative Genomics Viewer) or Circos can help in understanding complex data sets through graphical representations. These steps are critical for converting raw sequence data into actionable insights, supporting wide range of applications in genomics, from disease research to evolutionary biology. The efficiency and accuracy of NGS data analysis have revolutionized the field, enabling unprecedented advances in our understanding of genetic information.

Applications: There are vast applications of next generation sequencing method in different biological fields as shown in the below Figure. 2.



Source: (Khttps://www.drugtargetreview.com/article/32972/versatile-applications-of-next-generation-sequencing-in-pharmaceutics/).

Functional genomics: The approach of Illumina/Solexa method is sequencing by synthesis with reversible terminators. This read 30-40 base pairs of length which consists of 1Gb base pairs per run (17). Functional genomics often employs the Illumina/Solexa method, a technique known for its sequencing by synthesis approach using reversible terminators. This method allows for the accurate reading of DNA sequences by incorporating fluorescently labeled nucleotides one at a time. Each nucleotide addition is followed by a detection step, where the fluorescent signal is captured to determine the incorporated base. Once detected, the terminator is removed, allowing the next nucleotide to be added. This cycle repeats, enabling the sequencing of 30-40 base pairs in length. Remarkably, the Illumina/Solexa method can generate approximately 1Gb of sequence data per run, making it a powerful tool for high-throughput genomic analysis and a cornerstone of modern functional genomics.

Diagnostic virology: For the identification and discovery of numerous novel viruses as well as detection of various viral pathogens-leads to spreading of infection in pandemic (18). The influenza virus is identified by applying Illumina platform and respiratory tract infections, entero virus 109 are also detected by this method (19-25). Next generation sequencing tool also used for detection of tumor viruses like (HBV) hepatitis B virus, Human papilloma viruses 18 and 16, Ebola virus and human Spuma virus (26, 27). Merkel cell carcinoma is a dangerous human skin cancer which is pain less and caused by Merkel cell polyoma virus. HPyV9 is a strain of Merkel polyoma virus detected recently. The genome sequencing of retro virus and HIV are analyzed through deep sequencing technology (28-35).

Polyploid genetics: Next generation sequencing used to study the alloploidy. Tragopogon (Compositae) and Nicotiana (Solanaceae) have application of NGS for the investigation of genomic changes in polyploids (36). Since rice and sorghum contains numerous base pairs it is easy to analysis of gene sequencing (37).

GeneBank collections: National and international genebanks are facing many challenges to maintain the genebanks due to numerous size of collections. In addition to this they are also facing many legal problems NGS can greatly upgrade efforts of genetic characterization in gene-banks (38-40). Gene-banks also navigate complex legal and ethical landscapes. Issues related to intellectual property rights, access and benefit-sharing (ABS) agreements, and compliance with international treaties such as the Convention on Biological Diversity (CBD) and the Nagoya protocol add layers of complexity to genebanks operations. Legal uncertainties can hinder the exchange of genetic material between countries and institutions, impacting collaborative research and breeding programs. However, they face several challenges in maintaining and managing these collections due to the sheer size and complexity of the genetic material they hold.

Challenges in GeneBank management

Volume and diversity of collections: The enormous size and diversity of collections in genebanks pose significant logistical and technical challenges. Managing thousands to millions of accessions, ensuring their viability, and preventing genetic erosion over time require substantial resources and advanced technologies.

Role of NGS: Technologies offer transformative potential to address these challenges and enhance the genetic characterization and management of gene bank collections.

Comprehensive genetic characterization: NGS enables the detailed genetic characterization of large collections at unprecedented speed and accuracy. By sequencing entire genomes or specific gene regions, researchers can obtain comprehensive genetic profiles of accessions. This information helps in identifying genetic diversity, assessing genetic integrity and detecting any genetic drift or contamination in the collections.

Improved data management: The high-throughput nature of NGS generates vast amounts of genetic data, which can be integrated into advanced databases and bioinformatics tools. This facilitates better organization, retrieval, and analysis of genetic information, supporting efficient management and utilization of Genebanks resources.

Enhanced breeding programs: Detailed genetic information obtained through NGS can inform breeding programs by identifying valuable traits and genetic markers associated with disease resistance, yield improvement, and environmental adaptation. This accelerates the development of new crop varieties and enhances the overall genetic improvement efforts.

Support for legal compliance: With precise genetic data, genebanks can more effectively navigate legal frameworks by providing accurate documentation of genetic resources. This supports compliance with international agreements and strengthens the case for equitable benefit-sharing arrangements.

Resource optimization: NGS can help prioritize and streamline conservation efforts by identifying redundant or duplicate accessions. This allows genebanks to optimize storage space and resources, focusing on maintaining genetically unique and valuable specimens.

Drug discovery: Finding and developing drugs In the process of finding new drugs, one of the most crucial phases is target identification. The ultimate goal is to evaluate the therapeutic benefits that could result from changing suggested drug targets in patients taking part in clinical trials. Various techniques have been employed to discover and validate possible drug targets. In this situation, NGS can be utilized to steer and propel the creation of innovative treatment solutions that impact the biological targets linked to the advancement of illness. By offering thorough genome sequence information, NGS also offers a potent platform for discovering mutations linked to disease, making it possible to diagnose patients' illnesses with greater accuracy. That the appropriate therapy approach can be identified.

Advantages: It is a cost effective method and creating most accessible genome sequencing for different applications. Commonly used method in sequencing of various base pairs (41-44). It offers highly rising sensitivity for the detection of rare sequences in large genomic samples.

Cost-effective: NGS has become a cost-effective method for DNA sequencing, significantly reducing the cost per base pair compared to traditional methods. This affordability makes genomic sequencing more accessible for a variety of applications.

Wide application: NGS is commonly used for sequencing various base pairs, making it a versatile tool in genomics. It can handle a wide range of sequencing projects, from small-scale studies to large-scale genomic analyses.

High sensitivity: One of the key advantages of NGS is its high sensitivity. It can detect rare sequences in large genomic samples, which is crucial for identifying mutations, rare variants, and low-frequency alleles in complex biological samples.

Epigenetic research: NGS is invaluable for analyzing different types of histone modifications and conducting epigenetic research. It provides detailed insights into how gene expression is regulated by epigenetic changes, aiding in the understanding of gene function and disease mechanisms. These advantages underscore the transformative impact of NGS in advancing genomic research, enabling new discoveries, and improving our understanding of complex biological systems.

Disadvantages: Based on the applicable methodology in sequencing technique has some disadvantages like challenges has related to instrumentation, error rates, complexity of data analysis, sample quality requirements, ethical and privacy concerns (49-54).

It is time taken process, special knowledge is needs to generate accurate sequencing data, the cost of NGS technique is $100000. When compared to the Sanger sequencing it is high cost in many labs errors of sequencing may occur (55-59). NGS, while revolutionary, is also a time-consuming process that demands specialized knowledge to generate accurate sequencing data. The complexity of the technology requires expertise in both the operation of sequencing platforms and the analysis of the resulting data. The cost of implementing NGS, around $100,000 for a single setup, is significantly higher than that of traditional Sanger sequencing, making it a

substantial investment for many laboratories. Furthermore, despite its advanced capabilities, NGS is not immune to sequencing errors. These errors can arise from various sources, including sample preparation, sequencing chemistry and data interpretation, potentially impacting the reliability of the results. This necessitates rigorous quality control and validation procedures to ensure the accuracy of the sequencing data. Despite these challenges, the high-throughput and comprehensive nature of NGS continues to make it an invaluable tool in genomic research, provided that its limitations are carefully managed.

Challenges: NGS offers significant advantages, including high throughput, scalability, and cost-effectiveness. However, challenges remain in data storage, analysis, and interpretation. Advanced bioinformatics tools and expertise are essential to manage and extract meaningful insights from the vast amount of data generated.

Summary: NGS technologies allow for the simultaneous sequencing of millions to billions of DNA fragments, vastly increasing the speed, accuracy, and affordability of sequencing. Key NGS platforms include Illumina, Ion Torrent, and Pacific Biosciences (PacBio), each utilizing distinct sequencing chemistries and producing different read lengths. Illumina technology, the most commonly used, employs reversible terminator chemistry to generate short but highly accurate reads. Ion Torrent, on the other hand, detects hydrogen ions released during nucleotide incorporation, while PacBio specializes in producing long reads, making it particularly useful for assembling complex genomes and identifying structural variations. The NGS workflow is structured around several critical steps. First, DNA is fragmented, and adapters are ligated to facilitate attachment to the sequencing platform. This is followed by the amplification of DNA fragments using PCR techniques to generate sufficient quantities of DNA for sequencing. Once the amplified library is loaded onto the sequencing platform, each DNA fragment is sequenced in parallel, generating a massive amount of data. The final step involves data analysis, where sophisticated bioinformatics tools align the sequenced reads to a reference genome, perform *de novo* genome assembly, and identify genetic variants. This comprehensive process allows researchers to conduct in-depth genomic analyses that were previously unattainable. The applications of NGS are diverse and span multiple fields of study. In genomics, NGS is employed for whole-genome sequencing, exome sequencing, and targeted sequencing, which are crucial for detecting genetic variants and studying genetic diseases. Transcriptomics benefits from RNA sequencing, which analyzes gene expression patterns and identifies novel transcripts.

In epigenomics, NGS is used to study DNA methylation and histone modifications, offering insights into gene regulation mechanisms. Metagenomics leverages NGS to sequence environmental DNA, enabling the study of microbial diversity and ecosystem functions. In clinical diagnostics, NGS plays a pivotal role in identifying genetic mutations associated with diseases, guiding personalized medicine, and detecting pathogens with high precision. The advantages of NGS are numerous. It provides unparalleled throughput, scalability, and cost-effectiveness, making it accessible for a wide range of research applications. However, the technology is not without its challenges. The vast amount of data generated by NGS requires significant storage capacity and advanced computational tools for analysis. Interpreting the results also demands a high level of expertise in bioinformatics and genomics. Despite these challenges, the continuous development of more efficient algorithms and data analysis pipelines is helping to address these issues, making NGS an indispensable tool in modern genomics research.

Conclusion: The NGS is a ground breakage sequencing method. It is a revolutionized genomic research and high through put and cost effective DNA sequencing. It has many applications in various fields like -medicine, agriculture and bioinformatics. There are many challenges as data management and ethical considerations should be addressed for the utilization of next generation sequencing in the future genomics. Despite its transformative potential, NGS faces challenges such as data management, where the vast amounts of generated data require efficient storage, processing, and interpretation solutions.

References:

1. Behjati, S Patrick S Tarpey. Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridgeshire CB10 1SA, UK.

2. Tucker T, Marra M and Jan M Friedman. Massively parallel sequencing: the next big thing in genetic medicine. *Am. J. Hum. Genet*. 2009 Aug; 85(2):142-54.
doi: 10.1016/j.ajhg.2009.06.022.

3. Erwin L van Dijk, Yan Jaszczyszyn, Claude Themes Library preparation methods for next-generation sequencing: tone down the bias. *Exp Cell Res* 2014 Mar 10; 322(1):12-20. doi: 10.1016/j.yexcr.2014.01.008. Epub 2014 Jan 15.

4. Mardis ER A decades perspective on DNA sequencing technology. *Nature* 2011; 470: 198-203. DOI: 10.1038/nature09796.

5. Taishan Hu, Nilesh Chitnis, Dimitri Monos, Anh Dinh. Next-generation sequencing technologies: An overview. *Hum Immunol*. 2021 Nov; 82(11):801-811.
doi: 10.1016/j.humimm.2021.02.012. Epub 2021 Mar 19.

6. Sophia Yohe and Bharat Thyagarajan. Review of Clinical Next-Generation Sequencing
Arch Pathol Lab Med (2017) 141 (11): 1544–1557. doi.org/10.5858/arpa.2016-0501-RA.
7. Berglund E.C, Kiialainen A, Syvänen A.C. Next-generation sequencing technologies and applications for human genetic history and forensics. *Invest Genet*, 2 (2011), p. 23.
8. Cullinane A.R, Vilboux T, O'Brien K., *et al.* Homozygosity mapping and whole-exome sequencing to detect SLC45A2 and G6PC3 mutations in a single patient with oculocutaneous albinism and neutropenia. *J Invest Dermatol*, 131 (2011), pp. 2017-2025.
9. Gogol-Döring A and Chen W, 2012. An overview of the analysis of next generation sequencing data. *Methods Mol Biol*, 802 (2012), pp. 249-257.
10. Lai-Cheong J.E and McGrath J.A. Next-generation diagnostics for inherited skin disorders. *J Invest Dermatol*, 131 (2011), pp. 1971-1973,
11. Metzker M.L. Sequencing technologies—the next generation, *Nat Rev Genet*, 11 (2010), pp. 31-46,
12. Quail M.A, Smith M, Coupland P, *et al.* A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC Genome*, 13 (2012), p. 341.
13. Rehm H.L. Disease-targeted sequencing: a cornerstone in the clinic. *Nat Rev Genet*, 14 (2013), pp. 295-300
14. Sanger F, Nicklen S, Coulson A.R. DNA sequencing with chain-terminating inhibitors
Proc Natl Acad Sci, 74 (1977), pp. 5463-5467.
15. Xuan, J, Yu Y, Qing T, *et al.* Next-generation sequencing in the clinic: promises and challenges. *Cancer Lett* (2012)
16. Ayman Grada and Kate Weinbrecht (2013). Next-Generation Sequencing: Methodology and Application. *J Invest Dermatol*.
17. Olena Morozova, Marco A. Marra BC Cancer Agency Genome Sciences Centre, Suite 100, 570 West 7th Avenue, Vancouver, BC V5Z 4S6, Canada.
18. Luisa Barzon, Enrico Lavezzo,Valentina Militello, Stefano Toppo and Giorgio Palù Applications of Next-Generation Sequencing Technologies to Diagnostic Virology. *Int J Mol Sci* 2011; 12 (11):7861-84. doi: 10.3390/ijms12117861. Epub 2011 Nov 14.
19.Yongfeng, H.; Fan, Y.; Jie, D.; Jian, Y.; Ting, Z.; Lilian, S.; Jin, Q. Direct pathogen detection from swab samples using a new high-throughput sequencing technology. *Clin. Microbiol. Infect* 2011, *17*, 241–244.
20. Kuroda, M.; Katano, H.; Nakajima, N.; Tobiume, M.; Ainai, A.; Sekizuka, T.; Hasegawa, H.; Tashiro, M.; Sasaki, Y.; Arakawa, Y.; *et al.* Characterization of quasi-species of pandemic 2009 influenza A virus (A/H1N1/2009) by *de novo* sequencing using a next-generation DNA sequences. *PLoS One* 2010, 5.
21. Greninger, A.L.; Chen, E.C.; Sittler, T.; Scheinerman, A.; Roubinian, N.; Yu, G.; Kim, E.; Pillai, D.R.; Guyard, C.; Mazzulli, T.; *et al.* A metagenomic analysis of pandemic influenza A (2009 H1N1) infection in patients from North America. *PLoS One* 2010, 5.
22.Yang, J.; Yang, F.; Ren, L.; Xiong, Z.; Wu, Z.; Dong, J.; Sun, L.; Zhang, T.; Hu, Y.; Du, J.; *et al.* Unbiased parallel detection of viral pathogens in clinical samples using a metagenomic approach. *J. Clin. Microbiol* 2011, *49*, 3463–3469.
23.Yozwiak, N.L.; Skewes-Cox, P.; Gordon, A.; Saborio, S.; Kuan, G.; Balmaseda, A.; Ganem, D.; Harris, E.; DeRisi, J.L. Human entero virus 109: A novel interspecies recombinant entero virus isolated from a case of acute pediatric respiratory illness in Nicaragua. *J. Virol* **2010**, *84*, 9047–9058.
24.Cheval, J.; Sauvage, V.; Frangeul, L.; Dacheux, L.; Guigon, G.; Dumey, N.; Pariente, K.; Rousseaux, C.; Dorange, F.; Berthet, N.; *et al.* Evaluation of high throughput sequencing for identifying known and unknown viruses in biological samples. *J. Clin. Microbiol* 2011, *49*, 3268–3275.
25. Bishop-Lilly, K.A.; Turell, M.J.; Willner, K.M.; Butani, A.; Nolan, N.M.; Lentz, S.M.; Akmal, A.; Mateczun, A.; Brahmbhatt, T.N.; Sozhamannan, S.; *et al.* Arbo virus detection in insect vectors by rapid, high-throughput pyrosequencing. *PLoS Negl. Trop. Dis* 2010, *4*, e878:1–e878:12.
26. Weber, G.; Shendure, J.; Tanenbaum, D.M.; Church, G.M.; Meyerson, M. Identification of foreign gene sequences by transcript filtering against the human genome. *Nat. Genet* 2002, *30*, 141–142.
27. Xu, Y.; Stange-Thomann, N.; Weber, G.; Bo, R.; Dodge, S.; David, R.G.; Foley, K.; Beheshti, J.; Harris, N.L.; Birren, B.; *et al.* Pathogen discovery from human tissue by sequence-based computational subtraction. *Genomics* 2003, *81*, 329–335.
28. Feng, H.; Shuda, M.; Chang, Y.; Moore, P.S. Clonal integration of a polyomavirus in human Merkel cell carcinoma. *Science* 2008, *319*, 1096–1100.
29. Sauvage, V.; Foulongne, V.; Cheval, J.; Ar Gouilh, M.; Pariente, K.; Dereure, O.; Manuguerra, J.C.; Richardson, J.; Lecuit, M.; Burguiere, A.; Caro, V.; Eloit, M. Human polyomavirus related to African green monkey lymphotropic polyomavirus. *Emerg. Infect. Dis* 2011, *17*, 1364–1370.

30. Scuda, N.; Hofmann, J.; Calvignac-Spencer, S.; Ruprecht, K.; Liman, P.; Kuhn, J.; Hengel, H.; Ehlers, B. A novel human polyomavirus closely related to the African green monkey-derived lymphotropic polyomavirus. *J. Virol* 2011, *85*, 4586–4590.
31. Hacein-Bey-Abina, S.; Garrigue, A.; Wang, G.P.; Soulier, J.; Lim, A.; Morillon, E.; Clappier, E.; Caccavelli, L.; Delabesse, E.; Beldjord, K.; *et al.* Insertional oncogenesis in 4 patients after retrovirus-mediated gene therapy of SCID-X1. *J. Clin. Invest* 2008, *118*, 3132–3142.
32. Howe, S.J.; Mansour, M.R.; Schwarzwaelder, K.; Bartholomae, C.; Hubank, M.; Kempski, H.; Brugman, M.H.; Pike-Overzet, K.; Chatters, S.J.; de Ridder, D.; *et al.* Insertional mutagenesis combined with acquired somatic mutations causes leukemogenesis following gene therapy of SCID-X1 patients. *J. Clin. Invest* 2008, *118*, 3143–3150.
33. Bushman, F.; Lewinski, M.; Ciuffi, A.; Barr, S.; Leipzig, J.; Hannenhalli, S.; Hoffmann, C. Genome-wide analysis of retroviral DNA integration. *Nat. Rev. Microbiol* 2005, *3*, 848–858.
34.Pennisi, E. 2010. Semiconductors inspire new sequencing technologies. *Science* 327: 1190.
35. R Development Core Team (2010). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.
36. Ronaghi, M., M. Uhlén, and P. Nyrén. 1998. A sequencing method based on real-time pyrophosphate. *Science* 281: 363–365.
37. Buggs, R. J. A., S. Renny-Byfield, M. Chester, I. E. Jordon-Thaden, L. F. Viccini, S. Chamala, and A. R. Leitch el al. 2012. Next-generation sequencing and genome evolution in allopolyploids. *American Journal of Botany* 99: 372–382.
38. Berkman, P. J., K. Lai, M. T. Lorenc, and D. Edwards. 2012. Next-generation sequencing applications for wheat crop improvement. *American Journal of Botany* 99: 365–371.
39. Ashley N. Egan, Jessica Schlueter, David M. Spooner First published: 01 February 2012 doi.org/10.3732/ajb.1200020
40. FAO. 2010. *The second report on the state of the world's plant genetic resources for food and agriculture*. Food and Agriculture Organization, Rome, Italy.
41. McCouch, S. R., K. L. McNally, W. Wang, and R. S. Hamilton. 2012. Genomics of gene banks: A case study in rice. *American Journal of Botany* 99: 407–423.
42. Şule Ari and Muzaffer Arikan Next-Generation Sequencing: Advantages, Disadvantages, and Future. Chapter First Online: 24 August 2016. DOI: 10.1007/978-3-319-31703-8_5
43. Bahassi EM, Stambrook PJ (2014) Next-generation sequencing technologies: breaking the sound barrier of human genetics. *Mutagenesis* 29(5):303–310.
44. Bao S, Jiang R, Kwan W, Wang B, Ma X, Song YQ (2011) Evaluation of next-generation sequencing software in mapping and assembly. *J Hum Genet* 56:406–414.
45. Paul J. Hurd, Christopher J. Nelson Author Notes *Briefings in Functional Genomics*, Volume 8, Issue 3, May 2009, 174–183. doi.org/10.1093/bfgp/elp013.
46. Shendure' Next-generation DNA sequencing, *Nat Biotechnol*, 2008, vol. 26(pg. 1135-45).
47. Biggin MD, Gibson TJ, Hong GF (1983) Buffer gradient gels and 35S label as an aid to rapid DNA sequence determination. *Proc Natl Acad Sci* 80(13):3963–3965.
48. Bowers J, Mitchell J, Beer E, Buzby PR, Causey M, Efcavitch JW, Jarosz M, Krzymanska-Olejnik E, Kung L, Lipson D, Lowman GM, Marappan S, McInerney P, Platt A, Roy A, Siddiqi SM, Steinmann K, Thompson JF (2009) Virtual terminator nucleotides for next-generation DNA sequencing. *Nat Methods* 6(8):593–595.
49. Buermans HPJ, den Dunnen JT (2014) Next generation sequencing technology: advances and applications. *Biochim Biophys Acta* 1842(10):1932–1941
50. Chen EY (1994). The efficiency of automated DNA sequencing. In: Adams MD, Fields C, Venter JC (Eds) Automated DNA sequencing and analysis. Academic, San Diego, pp 3–9.
51. Chen CY (2014) DNA polymerases drive DNA sequencing-by-synthesis technologies: both past and present. *Front Microbiol* 5:305.
52. Chen F, Dong M, Ge M, Zhu L, Ren L, Liu G, Mu R (2013) The history and advances of reversible terminators used in new generations of sequencing technology. *Genomics Proteomics Bioinformatics* 11(1):34–40.
53. Crick FHC (1958). On protein synthesis. *Symp Soc Exp Biol* 7:138–163.
54. Derrington IM, Butler TZ, Collins MD, Manrao E, Pavlenok M, Niederweis M, Gundlach JH (2010) Nanopore DNA sequencing with MspA. *Proc Natl Acad Sci* 107(37):16060–16065
55. Schirmer, M., Ijaz, U. Z., D'Amore, R., Hall, N., & Quince, C. (2015). Next-generation sequencing technologies and their applications. Annual Review of Genomics and Human Genetics, 16, 223-244. https://doi.org/10.1146/annurev-genom-090413-025324.

56. M.A. Quail, M. Smith, P. Coupland, *et al.* A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers *BMC Genom*, 13 (2012), p. 341.
57. Rehm H.L. Disease-targeted sequencing: a cornerstone in the clinic. *Nat Rev Genet*, 14 (2013), pp. 295-300.
58. Sanger F, Nicklen S, and Coulson A.R. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci*, 74 (1977), pp. 5463-5467.
59. Xuan J, Yu Y, Qing T., Guo L and Shi L. Next-generation sequencing in the clinic: promises and challenges. *Cancer Lett* (2012). e-pub ahead of print 19 November 2012.