Int.J.Curr.Microbiol.App.Sci (2021) *10*(08): 329-333

International Journal of Current Microbiology and Applied Sciences
***ISSN: 2319-7706* Volume 10 Number 08 (2021)**
Journal homepage: http://www.ijcmas.com

Review Article https://doi.org/10.20546/ijcmas.2021.1008.039

Next Generation Sequencing: Principle and Applications

Arul Pandiyan* and Mahesh Dahake

Department of Veterinary Pathology, Nagpur Veterinary College, Maharashtra Animal and Fishery Sciences University, Nagpur-440006, India

**Corresponding author*

Keywords

Next Generation Sequencing, Clinical diagnostics, Sanger sequencing, Metagenomics

Article Info

Accepted:
15 July 2021
Available Online:
10 August 2021

A B S T R A C T

Next Generation Sequencing (NGS) also named as massively parallel sequencing, is a powerful new tool that can be used for the complex diagnosis and intensive monitoring of infectious diseases in veterinary medicine. NGS technologies are also being increasingly used to study the aetiology, genomics, evolution and epidemiology of infectious disease, as well as host pathogen interaction, host immune response including responses to antimicrobial treatment and vaccination. NGS approaches can be used as primary tools for the study of disease outbreaks by the identification and follow-up of transmission routes, thereby helping in the identification of outbreak origins in zoonotic diseases. The above applications of next generation sequencing in disease and cancer diagnosis, genome characterization and viral diversity, viral metagenomics and host pathogen interaction study make it a powerful tool for detailed analysis of complex populations in a minimalistic period of time.

Introduction

Histopathology is a technique used to directly visualize the diseased tissue for the collection of data which can be used for patient management. A variety of adjunctive tests have been developed over the years to identify the specific processes occurring within tissue samples. These tests range from simple histochemistry to immunohistochemistry and, more recently, to molecular testing of nucleic acids.

The structure of DNA was identified by Watson and Crick based on the fundamental DNA crystallography and X-ray diffraction work of Rosalind Franklin (Watson J.D.*et al.*, 1953). However, the first molecule to be sequenced was actually RNA – tRNA by Robert Holley (Holley R. W. *et al.*, 1965). Various research groups then began adapting these methods for DNA sequencing. Fredrick Sanger and colleagues, developed the first the chain-termination method and by 1986, the first automated DNA sequencing method had

been developed (Smith *et al.,* 1986). With the immense technological advances and the chain termination method, also known as Sanger sequencing, the human genome project was completed in 2003 (Abdellah, *et al.,* 2004).

In 2005, the first commercially available NGS platform was introduced, with which a plethora of genome sequencing projects that took many years with Sanger sequencing methods could now be completed within few hours with the NGS technology (Shendure, *et al.,* 2005).

The development of NGS has wide applicability in both clinical and research settings. It helps in the identification of germline or somatic mutations thereby helping in alleviating the clinical distress caused to the patient (Rizzo *et al.,* 2012). Its also a valuable tool in metagenomic research and for infectious disease diagnostics (Lefterova *et al.,* 2015). In 2020, NGS technology was also used in identifying and characterizing the SARS-CoV-2 genome for monitoring the COVID-19 pandemic (Mostafa *et al.,* 2020).

Next-Generation Sequencing Methods

The different sequencing methods in use for NGS technology are as described below.

Proton Detection Sequencing

The proton detection sequencing depends on counting hydrogen ions released during the polymerization of DNA. The pH changes are then detected by semiconductor sensor chips and converted to digital information (Rothberg *et al.,* 2011).

Pyrosequencing

Pyrosequencing uses the detection of pyrophosphate generation and light release to understand whether a specific base has been incorporated in a DNA chain (Ronaghi *et al.,* 1996).

Illumina Semiconductor Sequencing Method (SBS)

The SBS method is a reversible terminator sequencing method which works on the principle of ‘’bridge-amplification’’. During the synthesis reactions, the fragments bind to oligonucleotides on the flow cell, creating a bridge from one side of the sequence (P5 oligo on flow cell) to the other (P7), which is then amplified.

The added fluorescently-labeled nucleotides are detected using direct imaging (Buermans *et al.,* 2014).

Sequencing by Ligation Method

Sequencing by ligation method does not use DNA polymerase to create a second strand. The sensitivity of DNA ligase to base-pairing mismatches is utilized and the fluorescence produced used to determine the target sequence. These digital images taken after each reaction are further used for analysis.

DNA Nanoball Sequencing

DNA nanoball sequencing is a form of sequencing by ligation which utilizes rolling circle replication. Concatenated DNA copies are compacted into DNA nanoballs and bound to sequencing slides in a dense grid of spots which are ready for ligation-based sequencing reactions (Porreca *et al.,* 2010).

The NGS Workflow

Sample preparation (pre-processing)

Nucleic acids (DNA or RNA) are extracted from the blood, sputum, bone marrow samples.

Library preparation

Random fragmentation of the cDNA or DNA by enzymatic treatment or sonication is performed. The optimal fragment length depends on the platform that is being used. The fragments obtained are then end-repaired and further ligated to smaller generic DNA fragments called adapters. Adapters have defined lengths with known oligomer sequences to be compatible with the applied sequencing platform and identifiable where multiplex sequencing is performed. Multiplex sequencing by using individual adapter sequences per sample, helps in large numbers of libraries to be pooled and sequenced simultaneously in a single run. This pool of DNA fragments along with adapters attached is known as a sequencing library.

Size selection may then be performed, by gel electrophoresis or using magnetic beads, to remove any fragments that are too short or too long for optimal performance on the sequencing platform and protocol selected.

Library enrichment/amplification is then achieved using PCR. In techniques involving emulsion PCR, each fragment is bound to a single emulsion bead which will form the basis of sequencing clusters. Amplification is often followed by a cleaning step using magnetic beads to remove unwanted fragments and noise and also improve sequencing efficiency.

Sequencing

Depending on the selected platform and chemistry, clonal amplification of library fragments may occur prior to sequencer loading (emulsion PCR that permits amplification of DNA molecules in physically separated picoliter-volume water-in-oil droplets which further avoids formation of unproductive chimeras and other artifacts between similar DNA sequences) or on the sequencer itself (bridge PCR). Sequences are then detected and reported according to the platform selected (Head *et al.,* 2014).

Data analysis

The generated data files are analyzed depending on the workflow used. Analysis methods are highly dependent on the aim of the study.

A variety of different software packages are available which enable the data to be analysed. The main output of the data is usually in the form of a FASTQ file which contains the raw sequence and information about the quality of the sequence. Information about sequence quality is denoted by the "Phred" score which is provided to each base. This value indicates the probability of a base having been accurately called and therefore indicates the confidence with which a variant has been accepted as "true." (Hartman *et al.,* 2019).

In contrast to Sanger sequencing, which may have read lengths of up to 1 kb, NGS platforms produce short reads (typically 25–500 bases). Hence, once poor-sequence reads have been removed, the next step is to align the sequence fragments together. This process is known as sequence assembly. The sequence will be mapped and aligned automatically by the bioinformatics tools resulting in BAM (binary alignment/map) files.

These contain information on the sequence and its location in relation to the reference sequence. The BAM files can then be analysed further using software such as the Integrated Genome Viewer. The software will also produce a variant call file (VCF) which will contain information about the variant detected, the location of the variant and the number of reads at that location containing wild-type and variant sequences (Ronaghi *et al.,* 1996).

Applications of NGS

NGS technology has become an invaluable tool in both research and clinical diagnostic settings. It helps us study not only the genes and their subsequent involvement in disease causation butalso processes of mutation, rearrangement and fusion are identified. It is also used for the surveillance of antimicrobial resistance (Collineau *et al.,* 2019).

It is an important clinical tool in the diagnoses of certain diseases like cancer and other genetic diseases. It helps to detect familial cancer mutation carriers and to fully sequence all types of mutations for a large number of genes (hundreds to thousands) in a single test at a relatively low cost. In addition, it can also be used for RNA analysis. This enables the genomes of RNA viruses like influenza and SARS to be determined (Ozsolak *et al.,* 2011). Also, NGS helps in the detection and identification of various infectious agents involved in complex diseases such as post-weaning multisystemic wasting syndrome and the detection of infection agents in mixed infections in enteric disease complexes.

NGS approaches have been of use in routine diagnostics to monitor the genomic diversity of AIV, early emergences and transmission of these viruses from waterfowl to domestic poultry and also to study of genomic recombination, which has an important role in bacterial and viral evolution.

Epigenomic sequencing helps in studying the changes caused by histone modifications and DNA methylation. There are different methods employed for the study of epigenetic mechanisms, including whole genome bisulfate sequencing (WGBS), chromatin immunoprecipitation (ChIP-seq) and methylation dependent immunoprecipitation (MeDIP-seq) followed by sequencing (Ku *et al.,* 2011). Depending on the selected method, the complete DNA methylome and histone modification profiles can be mapped and studied, giving information regarding the genomic regulatory mechanisms.

Metagenomic sequencing can provide information for samples collected in a specific environment. It provides the comparison of differences and interactions between mixed microbial populations and host responses. Some of the potential applications of metagenomic sequencing include, but are not limited to, infectious disease diagnostics and infection surveillance, antimicrobial resistance monitoring, microbiome studies and pathogen discovery (Chiu *et al.,* 2019)

Considering the various merits of next generation sequencing in comparison with other techniques like histopathology and immunohistochemistry, next generation sequencing has emerged to be the supreme diagnostic tool for disease surveillance. It allows rapid assessment of large number of cells, ability to analyse many samples quickly, ability to reanalyse data and high accuracy and reproducibility. Due to its great potential, the use of next generation sequencing has been expanded to diverse fields of biological sciences and various health areas and is routinely used in clinical diagnostics, biotechnology and basic and applied research (Errante *et al.,* 2016).

References

Abdellah Z, Ahmadi A, Ahmed S. (2004). Finishing the euchromatic sequence of the human genome. Nature, 431, 931-945.

Buermans H P J, den Dunnen J T. (2014). Next generation sequencing technology: Advances and applications. BiochimBiophysActa - Mol Basis Dis., 1842(10), 1932-1941.

Collineau L, Boerlin P, Carson C A. (2019).

Integrating whole-genome sequencing data into quantitative risk assessment of foodborne antimicrobial resistance: A review of opportunities and challenges. Front Microbiol., 10, 1107.

Chiu C Y, Miller S A. (2019). Clinical metagenomics. Nat Rev Genet., 20(6), 341-355.

Holley R W, Apgar J, Everett G A. (1965). Structure of a ribonucleicacid. Science, 147, 1462-1465.

Head S R, Kiyomi Komori H, LaMere S A. (2014). Library construction for next-generation sequencing: Overviews and challenges. Biotechniques., 56(2), 61-77.

Hartman P, Beckman K, Silverstein K. (2019). Next generation sequencing for clinical diagnostics: Five year experience of an academic laboratory. Mol Genet Metab Reports., 19, 100464.

Ku C S, Naidoo N, Wu M, Soong R. (2011). Studying the epigenome using next generation sequencing. J Med Genet., 48(11), 721-730.

Lefterova M I, Suarez C J, Banaei N, Pinsky B A. (2015). Next-Generation Sequencing for Infectious Disease Diagnosis and Management: A Report of the Association for Molecular Pathology. J Mol Diagnostics,17(6), 623-634.

Mostafa H H, Fissel J A, Fanelli B. (2020). Metagenomic next-generation sequencing of nasopharyngeal specimens collected from confirmed and suspect covid-19 patients. MBio., 11(6), 1-13.

Ozsolak F, Milos P M. (2011). RNA sequencing: Advances, challenges and opportunities. Nat Rev Genet., 12(2), 87-98.

Porreca G J. (2010). Genome sequencing on nanoballs. Nat Biotechnol., 28(1), 43-44.

Rizzo J M, Buck M J. (2012). Key principles and clinical applications of "next-generation" DNA sequencing. Cancer Prev Res., 5, 887-900.

Rothberg J M, Hinz W, Rearick T M. (2011). An integrated semiconductor device enabling non-optical genome sequencing. Nature, 475, 348-352.

Ronaghi M, Karamohamed S, Pettersson B, Uhlén M, Nyrén P. (1996). Real-time DNA sequencing using detection of pyrophosphate release. Anal Biochem., 242, 84-89.

Smith L M, Sanders J Z, Kaiser R J. (1986). Fluorescence detection in automated DNA sequence analysis. Nature, 321, 674-679.

Shendure J, Porreca G J, Reppas N B. (2005). Molecular biology: Accurate multiplex colony sequencing of an evolved bacterial genome. Science, 309, 1728-1732.

Watson J D, Crick F H C. (1953). Molecular structure of nucleic acids: A structure for deoxyribose nucleic acid. Nature, 171, 737-738.

How to cite this article:

Arul Pandiyan and Mahesh Dahake. 2021. Next Generation Sequencing: Principle and Applications. *Int.J.Curr.Microbiol.App.Sci.* 10(08): 329-333.
doi: https://doi.org/10.20546/ijcmas.2021.1008.039