Acta Scientific Veterinary Sciences (ISSN: 2582-3183)

Volume 2 Issue 10 October 2020

Review Article

A Brief Review of the Next Generation Sequencing

Afnan Saleem*

PhD Scholar, Division of Animal Biotechnology, Faculty of Veterinary Sciences and Animal Husbandry, SKUAST-Kashmir, India

***Corresponding Author:** Afnan Saleem, PhD Scholar, Division of Animal Biotechnology, Faculty of Veterinary Sciences and Animal Husbandry, SKUAST-Kashmir, India.

Received: August 26, 2020

Published: September 16, 2020

Abstract

Since the discovery of DNA in 1869, the development of classical genetics and research of how our bodies work have all led us to this point in human history where we are competent enough to completely sequence all the genes in a living body. The nature of any biological inquiry has now been transformed by the high throughput sequencing technologies to generate genome scale data sets. Based on Sanger's founding discovery of DNA sequencing, next generation DNA sequencing platforms are rapidly evolving to exemplify an elegant interplay between biology and technology. Over the past few years, second and third generation sequencing technology has dramatically accelerated the comprehensive analysis of biological and biomedical research. This review enlists a short summary of second and third generation sequencing technologies with an overview of the advances in the sequencing platforms.

Keywords: Next Generation Sequencing; Platforms; Second Generation; Third Generation; Applications

Abbreviations

NGS: Next Generation Sequencing; SOLiD: Sequencing by Oligonucleotide Ligation and Detection; SMRT: Single Molecule Real Time Sequencing; ZMW: Zero Mode Waveguides; PCR: Polymerase Chain Reaction

Introduction

In 1988, the Human Genome Project began and after its completion in 2003, second and third generation sequencing technologies arrived on the market. The so-called "next-generation" sequencing (NGS) technologies allows us, in a short time, low cost and in parallel, to sequence millions to billions of DNA nucleotides, minimizing the need for the fragment-cloning methods and overcoming the limitations of the original Sanger sequencing methods used to sequence the first human genome. Using different sequencing technologies, a wide variety of different NGS platforms have been made available. Sequencing of millions of small fragments of DNA in parallel has revolutionized the sequencing platforms. Further, bioinformatics analyses map the individual reads to the reference genome. In 1977, the Sanger method/chain termination method [1] of gene sequencing was developed. Despite its low output and accuracy, several new methods have been directly developed from the advances made from the Sanger's method. Second generation sequencing platforms require amplified sequencing libraries making it time consuming and costly while the third generation single molecule sequencing can be undertaken without the requirement of amplified DNA clones.

Massive-parallel and Next-generation DNA sequencing are used to refer collectively the available high throughput DNA sequencing technologies. These sequencing technologies supervise the sequential addition of nucleotides to spatially assorted DNA templates. However, the difference in the sequencing platforms lies in how the templates are generated and how they are detected to expose their sequences.

Second generation sequencing

Basic characteristics of the second generation sequencing involve the shotgun sequencing of the fragmented genomic DNA or cDNA ligated to the linkers or adapters for construction of template libraries. A solid surface or beads are used for the library amplification and the incorporated nucleotide is monitored by the luminescence generation or changes in electrical current. A major defining feature of the second generation sequencing is the parallelization of millions of nucleotide short reads in a much shorter time. Second generation sequencing includes Roche 454, Illumina, SOLiD and Ion Torrent.

Roche 454 pyrosequencing

Roche 454 pyrosequencing by synthesis [2] was the first successful second generation sequencing system developed in 2005. This technology made use of sequencing chemistry, wherein the visible light produced by ATP sulfurylase, luciferase, DNA polymerase enzymatic system and is measured in proportion to the amount of released pyrophosphate in the newly synthesized DNA chain [3]. This ensures that the reaction can be detected at a given light intensity and can analyze a large number of samples in parallel, vastly improving the output as compared to the original Sanger method. Roche in 2008 released the upgraded 454 GS FLX Titanium system with an average read length to 700 bp, 99.997% accuracy and 0.7 Gb of data per run output within 24 hrs. The major drawbacks of Roche 454 in most cases was the high error rates in homopolymer repeats and the high reagent cost. This technology has been discontinued for several years now.

Illumina (Solexa) HiSeq and MiSeq

Illumina sequencer also adopted the sequencing by synthesis technology using removable fluorescently labeled chain-terminating nucleotides that are capable of producing a larger output at a lower reagent cost [4]. A process called bridge amplification or cluster generation occurs allowing generation of copies of a clonally enriched template DNA to be made into miniaturized colonies called as polonies [5]. The sequencing data output per run is approximately higher (600 Gb) with shorter read lengths (~100 bp) and cheaper cost cheaper cost. Illumina provides atleast eight industrial level sequencing machines (NextSeq 500, HiSeq series 2500, 3000 and 4000 and HiSeq X series five and ten) having mid to high output (120 - 1500 Gb). Also, MiSeq is a small compact laboratory sequencer using the same sequencing and polony technology with faster turnover rates and an output of 0.3 to 15 Gb [6]. Further, Illumina's new method of TruSeq technology involving synthetic long reads resolves the complex, repetitive transposable elements and improves *de novo* assembly [7].

Sequencing by oligonucleotide ligation and detection (SOLiD)

SOLiD [8] (sequencing by oligonucleotide ligation and detection) systems was first released by Applied Biosystems Instruments (ABI) in 2008. Based on a 2-nucleotide sequencing by ligation (SBL), it involves sequential annealing and subsequent ligation of probes to the template [3,5]. Previously, SOLiD 5500 W series were used for projects involving exomes, transcriptomes and whole genomes [5]. Its sequencer used fluorescently labeled octamer probes in repeated annealing and ligation cycles which were eventually cross examined and decoded using Exact Call Chemistry [9]. Since each base was interrogated twice, it conferred a major advantage to the platform. However, shorter read lengths (50 - 75 bp), long run times and the need of expert computational infrastructure were the major drawbacks. This service has also been discontinued since 2016.

Ion torrent

Ion Torrent system was introduced by the 454 sequencing inventors with two major changes [10]. Instead of a light signal being detected by an optical system, the pH change in the surrounding solution is detected electronically proportional to the number of nucleotides incorporated. Secondly, electronic sensors are placed at the bottom of the flow cells while the sequencing reactions are performed within a microchip amalgamated with flow cells. Sequencers in the market that use Ion Torrent technology are the Ion Personal Genome Machine (PGM), a bench-top sequencer with 11.1 million sensors and the high-throughput Proton sequencer with more than 165 million sensors. Ion Torrent chip has been engineered to detect individual protons by an ion-sensitive field-effect transistor sensor. Chip when placed inside the flow cell is flushed with unlabelled dNTPs in the presence of DNA polymerase. H proton is released owing to the incorporation of the nucleotide in the DNA chain and thus, a pH change is detected. Problems in reading homopolymer repeats are the major disadvantages while the relative longer read length, cheaper price, less turnaround time are the major advantages [11].

Third generation sequencing

Third generation single molecule sequencing actually allows for sequencing without PCR amplification of the DNA enabling shorter preparation times, while decreasing the chances of error. This allows for billions of unique fragments to be independently

sequenced at the same time. Third Generation Sequencing includes Helicos True Single Molecule Sequencing, Single Molecule Real time Sequencing (SMRT) and Oxford Nanopore.

Helicos sequencing

Helicos sequencing provider Seqll sequences the genomic DNA and RNA by means of Helicos sequencing system and HeliScope single-molecule sequencers [5]. This method is an amalgamation of sequencing by hybridization and sequencing by synthesis using a DNA polymerase Sheared DNA is tailed with polyA and hybridized to a disposable glass flow cell surface surrounded by oligo-dT, allowing a parallel sequencing by synthesis of billions of molecules. The process of adding fluorescent nucleotides with a terminating nucleotide will pause the process until each nucleotide of the DNA sequence has been captured. This process repeats itself until all the fragments have been sequenced completely [12]. This process largely avoids size bias or GC content bias since there is no requirement of PCR amplification or ligation when compared with other methods [13]. Sequencing read lengths range from 25 to 60 bases.

Single-molecule real-time (SMRT) sequencing

Pacific Biosciences has developed the Single Molecule Real time DNA Sequencing (SMRT) and markets PacBio RS II sequencer [14]. SMRT sequencer uses Zero mode waveguides (ZMWs) which contain 1.5 lac ultra-microwells where each molecule of DNA polymerase is immobilized at the bottom of the well using biotin-streptavidin system. Once the coupling of the single stranded template DNA with the immobilized polymerase takes place, each nucleotide incorporation is detected by the addition of the fluorescently labeled dNTP analogs. ZMWs are continuously monitored using CCD cameras and a series of pulses are converted into single molecular traces corresponding to the template sequence. This platform of sequencing allows for a faster genome assembly than comparative technologies since all four nucleotides are added simultaneously and measured in real time. 99.3% accuracy has been reported with a read length of 900 bp [4].

Nanopore sequencing (MinION and PromethION)

In 2012, Oxford Nanopore technologies introduced latest single-molecule sequencing systems. MinION Mkl sequencing machine is a portable device of the size of a USB drive for DNA and RNA sequencing that can attach directly to a computer or a laptop while as PromethION system is a small bench top. Nanopore uses pores which are formed from proteins and are capable of detecting different DNA bases. The whole idea behind nanopores is that every single nucleotide will alter the ionic current as they pass through the pore generating time specific signals which are then evaluated in real time [15]. The shape of the molecule translocating through the pore will determine the flow of ion current [16]. The major advantage of this approach is the minimal sample preparation, no PCR amplification or ligation steps and longer read lengths (kbp range). However, optimizing the speed of DNA translocation through the nanopores seems the major problem. This will ensure reliable measurements of the current changes as well as reduce high error rates [15].

Conclusion

Over the last years, the advancements made on sequencing technology have been impressive. It has allowed the researchers to virtually ask any question related to the genome or transcriptome of any organism contributing to the genome wide knowledge of organisms by substantial cheaper, friendlier and more flexible high throughput sequencing methods. This technology has an high impact among the various areas of molecular biology, agriculture, bioindustry but its surely comes up with various risks involving ethical and regulatory risks in regard to privacy of genetic information, eugenics and the spectrum of bioterrorism. Also, many challenges still remain there in terms of huge data acquisition and its storage, data analysis and its interpretation. Undoubtedly, the future developments will rely on the newer technologies and collaborative efforts for high throughput data production and analysis.

Bibliography

1. Sanger F., *et al*. "DNA sequencing with chain-terminating inhibitors". *Proceedings of the National Academy of Sciences of the United States of America* 74 (1977): 5463-5467.

2. Wheeler D., *et al*. "The complete genome of an individual by massively parallel DNA sequencing". *Nature* 452 (2008): 872-876.

3. Margulies M., *et al*. "Genome sequencing in microfabricated high-density picolitre reactors". *Nature* 437 (2005): 376-380.

4. Metzker ML. "Sequencing technologies - The next generation". *Nature Reviews Genetics* 11 (2010): 31-46.

5. Shendure J and Ji H. "Next-generation DNA sequencing". *Nature Biotechnology* 26 (2008):1135-1145.

6. Liu L., *et al*. "Comparison of next-generation sequencing systems". *Journal of Biomedicine and Biotechnology* (2012): 251364.

7. McCoy RC., *et al*. "Illumina TruSeq synthetic long-reads empower de novo assembly and resolve complex, highly-repetitive transposable elements". *PLoS One* 9 (2014): e106689.

8. McKernan KJ., *et al*. "Sequence and Structural Variation in a Human Genome Uncovered by Short-Read, Massively Parallel Ligation Sequencing Using Two-Base Encoding". *Genome Research* 19 (2009): 1527-1541.

9. Guzvic M. "The history of DNA sequencing". *Journal of Medical Biochemistry* 32 (2013): 301-312.

10. Rothberg JM., *et al*. "An integrated semiconductor device enabling non-optical genome sequencing". *Nature* 475 (2011): 348-352.

11. Diekstra A., *et al*. "Translating Sanger-based routine DNA diagnostics into generic massive parallel ion semiconductor sequencing". *Clinical Chemistry* 61 (2015): 154-162.

12. Eid J., *et al*. "Real-time DNA sequencing from single polymerase molecules". *Science* 323 (2009): 133-138.

13. Hart C., *et al*. "Single-molecule sequencing: sequence methods to enable accurate quantitation". *Methods in Enzymology* 472 (2010): 407e430.

14. Schadt EE., *et al*. "A window into third-generation sequencing". *Hum Mol Gene* 19 (2010): R227-R240.

15. Bayley H. "Nanopore sequencing: From imagination to reality". *Clinical Chemistry* 61 (2015): 25-31.

16. Stoddart D., *et al*. "Single nucleotide discrimination in immobilized DNA oligonucleotides with a biological nanopore". *Proceedings of the National Academy of Sciences of the United States of America* 106 (2009): 7702-7707.