Journal of Pharmacognosy and Phytochemistry

Available online at www.phytojournal.com

E-ISSN: 2278-4136
P-ISSN: 2349-8234
JPP 2018; 7(6): 1316-1320
Received: 22-09-2018
Accepted: 24-10-2018

Nirja Thakur
Department of Biotechnology, YSP UHF, Nauni, Solan, Himachal Pradesh, India

Poonam Shirkot
Department of Biotechnology, YSP UHF, Nauni, Solan, Himachal Pradesh, India

Himanshu Pandey
Department of Biotechnology, YSP UHF, Nauni, Solan, Himachal Pradesh, India

Kanika Thakur Rachappanavar
Department of Seed Science and Technology, YSP UHF, Himachal Pradesh, India

Correspondence
Nirja Thakur

Next generation sequencing - Techniques and its applications

Nirja Thakur, Poonam Shirkot, Himanshu Pandey, Kanika Thakur, Vinay kumar and Rachappanavar

Abstract
First-generation sequencing was widely used during the 1980s and 1990s, but has been outdated in terms of output by NGS. These techniques are not time consuming and labour intensive as compare to first generation sequencing techniques. In contrast to Sanger sequencing, the speed of sequencing and amounts of DNA sequence data generated with NGS, which is considered a "high-throughput technology", are exponentially greater and are produced at significantly reduced costs. Third-generation sequencing uses parallel sequencing similar to NGS, but unlike NGS, third-generation sequencing uses single DNA molecules rather than amplified DNA as a template. Thus, third generation sequencing potentially eliminates errors in DNA sequence introduced in the laboratory during the DNA amplification process. The growing power and reducing cost sparked an enormous range of applications of Next generation sequencing (NGS) technology This review focuses on different techniques of next generation sequencing, their procedure on how they work as well as on their applications.

Keywords: Next generation sequencing techniques, Basic technology,

Introduction
NGS has revolutionized not only the genome sequencing and personal medicine, but it has also altered genome research which was previously done. Sequencing is execution of technique that helps to determine the number and order of nucleotides that are present in DNA of particular organism. Indispensible knowledge gained from DNA sequencing aid many biological research and other different kinds of fields like biotechnology, forensic science and biological systematic. The peculiar arrangement of bases significantly affect the health, e.g. How we response to particular disease and what kind of medication is important to cure that? Next-generation sequencing (NGS) is a type of DNA sequencing technique that utilize parallel sequencing of multiple small fragments of DNA to identify sequence (Rizzo *et al.* 2012) [20]. High-throughput technology is speedy and generates large amount of data which is exponentially greater as compare to Sanger technique moreover it is cost effective also (Voelkerding *et al*, 2009) [39]. Third-generation sequencing is quite similar to NGS, but in contrary to NGS, third-generation sequencing utilizes single DNA molecules instead of amplified DNA as a template. Thus, third generation sequencing possibly minimize the errors in DNA sequence risen in the laboratory during the DNA amplification process (Munroe *et al.* 2010; Schadt *et al.* 2010) [22, 32]

Sequencing Methods: can be classified into three basic groups

Classical Methods
- Maxam and Gilbert's chemical degradation method
- Sanger's method (Dideoxynucleotide chain termination method)

1. Maxam and Gilbert's chemical degradation method
Allan Maxam and Walter Gilbert brought out first DNA sequencing method in 1977 based on chemical alteration of DNA and subsequent cleavage at precised bases alternatively know as chemical sequencing, this method employ purified samples of double-stranded DNA which is to be required without further cloning. Maxam-Gilbert sequencing technique is based on radioactive labeling at one 5' end of the DNA and requires purification of the DNA fragment to be sequenced. Chemical treatment then creates a break at a small proportion of one or two of the four nucleotide bases in each of four reactions (G, A+G, C, C+T). The concentration of

the modifying chemicals kept such that it targets on average one modification per DNA molecule. Thus a series of labeled fragments are generated, from the radiolabeled end to the first "cut" site in each molecule. The fragments in the four reactions tubes are electrophoresed side by side in denaturing acrylamide gels for size separation further for visualization of the fragments, the gel is detected in X-ray film for autoradiography, yielding a series of dark bands each corresponding to a radiolabeled DNA fragment, from which the sequence may be inferred. This method's involve the use of radioactive labeling and its technical complexity discourage extensive use of this method but further refinement had been made in this technique which enable it to deploy in various research (Maxam &Gilbert, 1977) [19].

2. Sanger's method (Dideoxynucleotide chain termination method

The chain-termination method developed by Frederick Sanger and coworkers in 1977 soon became popular, owing to its relative simplicity and reliability (Sanger *et al.* 1977; Sanger & Coulson, 1975) [30, 31]. The chain-terminator method used fewer toxic chemicals and lower amounts of radioactivity in contrast with Maxam and Gilbert method. The Sanger method was soon automated because of its less complexity and was therefore used in the first generation of DNA sequencers. Sanger sequencing was popular method that prevailed from 1980s until the mid-2000.Great advances were made in this technique with the advent of new technologies, such as fluorescent labelling, capillary electrophoresis, and general automation. These developments allowed much more efficacy in field of sequencing, leading to cost effictiveness. The Sanger method, in mass production form, firstly used in human genome in 2001. However, later in the decade, different approaches replaced Sanger method, bringing down the cost per genome from $100 million in 2001 to $10,000 in 2011 (Wetterstrand, 2013.) [40]

Next Generation Sequencing Techniques

1. Second generation sequencing techniques
2. Third generation sequencing techniques
3. Fourth generation sequencing techniques

Second generation sequencing techniques

- 454 sequencing or pyrosequensing
- Illumina (Solexa) HiSeq and Mi Seq sequencing
- SOLID Sequencing
- Polony sequencing
- Massively parallel signature sequencing (MPSS)

454 sequencing or Pyro sequencing

The principle of Pyro sequencing was first reported in 1993 (Nyren *et al.*1993) using streptavidin coated magnetic beads with recombinant DNA polymerase lacking 3´to 5´exonuclease activity (proof-reading) and luminescence detection using the firefly luciferase enzyme (Nyren & Lundin, 1985) [23]. A mixture containing three enzymes (DNA polymerase, ATP sulfurylase and firefly luciferase) and a nucleotide (d NTP) are added to single stranded DNA to be sequenced and the incorporation of nucleotide is regulated by measuring the light emitted. The intensity of the light determining factor for nucleotides incorporated, thus showing how many complementary nucleotides are present on the template strand. The nucleotide mixture is withdrawn before the next nucleotide mixture is added. This process is performed again with each of the four nucleotides until the DNA sequence of the single stranded template is determined.

Illumina (Solexa) HiSeq and MiSeq sequencing

Illumina purchased the Solexa Genome Analyzer in 2006 and commercialized it in 2007 (Shendure *et al.*, 2015) [35]. Presently, it is considered the most successful sequencing system claiming >70% dominance of the market, particularly with the HiSeq and MiSeq platforms. Unlike Roche 454, the Illumina sequencer adopt the technology of sequencing by synthesis using removable fluorescently labeled chain-terminating nucleotides that are able to produce a larger cost effective output (Metzker, 2010; Margulies *et al.* 2005) [21, 18], The clonally enriched template DNA for sequencing is created by PCR bridge amplification (also known as cluster generation) into miniaturized colonies called polonies (Shendure *et al.*, 2008) [35]. The output of sequencing data per run is higher (600 Gb), the length of reads are shorter (approximately 100 bp), the cost is cheaper, and the run times are much longer (3-10 days) as compare to other systems (Liu *et al.* 2012) [16]. Illumina provides diverse range of sequencing machnine ranges from Next Seq 500, HiSeq series 2500, 3000, and 4000, and HiSeq X series five and ten with mid to high output (120–1500 Gb) as well as a compact laboratory sequencer called the MiSeq, which, although small in size having output of 0.3 to 15 Gb furthermore turnover rates appropriate for targeted sequencing of clinical and small laboratory applications [Illumina Sequencer Comparison Table [Internet]. Available from: http://www.illumina.com/systems/sequencing.html [Accessed: 2015-06-16].]. The MiSeq utilises the same sequencing and Polony technology such as the high-end machines, but it can produce sequencing results in 1 to 2 days (Liu *et al.* 2012) [16]. Illumina's latest method of synthetic long reads apply Tru Seq technology apparently enhance *de novo* assembly and resolves complex, highly repetitive transposable elements (McCoy *et al.*, 2014) [20]

SOLiD sequencing

Applied Biosystems' (now a Life Technologies brand) SOL iD technology uses sequencing by ligation. In this technique, a pool of all possible oligonucleotides of particular length are labeled according to the position of sequence. Oligonucleotides are annealed and ligated; the ligation is carried out by enzyme DNA ligase for matching sequences results in a signal informative of the nucleotide at that position. Before sequencing, the DNA is amplified by emulsion PCR. The resulting beads, each containing single copies of the same DNA molecule, are deposited on a glass slide (Valouev *et al.*, 2008) [38]. The result is sequences of quantities and lengths comparable to Illumina sequencing. Schuster SC (2008) [33]. This sequencing by ligation method has been reported to have some issue sequencing palindromic sequences (Huang *et al.*, 2012) [13].

Polony sequencing

The Polony sequencing method was developed at George M. Church laboratory in Harvard, was first among high-throughput sequencing systems and was applied for sequencing full genome e coli in 2005. It amalgamate an *in vitro* paired-tag library with emulsion PCR, an automated microscope, and ligation-based sequencing chemistry to sequence an *E. coli* genome at an accuracy of >99.9999% and a cost approximately 1/9 that of Sanger sequencing (Shendure *et al.*, 2005) [34]. The technology was licensed to Agen court

Biosciences, subsequently spun out into Agincourt Personal Genomics, and eventually included into the Applied Biosystems SOL iD platform. Applied Biosystems was later occupied by Life Technologies, now part of Thermo Fisher Scientific.

Massively parallel signature sequencing (MPSS)

A novel method of high-throughput sequencing technologies that is massively parallel signature sequencing (or MPSS), was developed in the 1990s at Lynx Therapeutics, a company founded in 1992 by Sydney Brenner and Sam Eletr. MPSS was a bead-based method that apply a complex approach of adapter ligation followed by adapter decoding, reading the sequence in increments of four nucleotides. This method made it liable to sequence-specific bias or loss of particular sequences. The complexity involve in this technology made MPSS performed only 'in-house' by Lynx Therapeutics and no DNA sequencing machines were sold to other laboratories. Lynx Therapeutics linked up with Solexa (later acquired by Illumina) in 2004, leading to the generation of sequencing-by-synthesis, a simpler approach acquired from Manteia Predictive Medicine, which rendered MPSS obsolete. The essential properties of the MPSS output were typical of later high-throughput data types, including hundreds of thousands of short DNA sequences. In the case of MPSS, these were utilized for sequencing cDNA for measurements of gene expression levels (Brenner *et al.*, 2000) [3].

Third generation sequencing techniques

- Single molecule real time sequencing Biosciences
- Heliscope Sequencing

Single molecule real time sequencing Biosciences

Pacific Biosciences involve in trading of PacBio RS II sequencer and the SMRT real-time sequencing system (Schadt *et al.*, 2010) [32] SMRT sequencing is executed in SMRT cells that accommodate 150,000 ultra-microwells at a zeptoliter scale where one molecule of DNA polymerase is immobilized at the bottom of each well with the help of biotin-streptavidin system in nanostructures known as zero-mode waveguides (ZMWs). Once the template single-strand DNA is integrated with immobilized DNA polymerase, fluorescently labeled dNTP analogs are incorporated and then nucleotides are added into the growing strand. CCD cameras continuously detect the 150,000 ZMWs as a series of distinguished pulses that are transformed into single molecular traces representing the template sequences. Since all four nucleotides are added simultaneously and measured in real time, the speed of sequencing is much rapid as compared to technologies where individual nucleotides are flushed sequentially moreover the accuracy was 99.3% initially with read lengths of about 900 bp (Metzker, 2010) [21], circularizing the template and sequencing it various times employing a technology called SMRT bell templates provided longer reads and achieved the accuracy to >99.999% (Travers *et al.*2010; Koren *et al.*, 2013) [37, 15]. Once sequencing commences, the system's computational Blade Center conducts real-time signal processing, base calling, and quality assessment. Primary analysis data, including trace and pulse data, read-length, distribution, polymerase speed, and quality measurement, is streamed directly to the secondary analysis software called SMRT Analysis that is capable of processing sequencing data in real time. The secondary analysis tools also comprise a full suite of tools to analyze single-molecule sequencing data for various applications.

Heliscope Sequencing

The Helicos sequencing system was the first commercial execution of single-molecule fluorescent sequencing (Shendure *et al.*, 2008; Thompson & Steinmann, 2010) [35], marketed by the now bankrupt Helicos Biosciences. Today, the sequencing provider Seqll sequences genomic DNA and RNA using the Helicos sequencing system and Heli Scope single-molecule sequencers. DNA is cleaved, tailed with poly A, and hybridized to a flow cell surface having oligo-dT for sequencing-bysynthesis of billions of molecules in parallel. The poly A-tailed fragments of DNA molecules are attached directly to the oligo-dT50 bound on the surface of disposable glass flow cells. The incorporation of fluorescent nucleotides with a terminating nucleotide discontinue the cyclical process until an image of one nucleotide for each DNA sequence has been captured, and then the process is repeated until the fragments have been completely sequenced (Eid *et al.*, 2009; Travers *et al.*, 2010) [8, 37]. This sequencing system is a combination of sequencing by hybridization and sequencing by synthesis using a DNA polymerase (Fuller *et al.*, 2009) [9]. Sample preparation does not involve ligation or PCR amplification and, therefore, largely rule out the GC content and size biases observed in other technologies (Hart *et al.*, 2010) [11]. The Heli Scope sequencing read lengths range from 25 to over 60 bases, with 35 bases being the average. This method has successfully employed in the human genome (Pushkarev *et al.*, 2009) [27] that impart disease signatures in a clinical evaluation (Ashley *et al.*, 2010) [1] & sequenced RNA to produce quantitative transcriptomes of tissues and cells (Hickman *et al.*, 2013) [12].

Fourth generation sequencing techniques

DNA nanoball sequencing: DNA nano ball sequencing is a type of high throughput sequencing technology that helps to determine the complete genomic sequence of an organism. The company Complete Genomics utilizes this technique to sequence samples submitted by different researchers. The method employ rolling circle replication to amplify small fragments of genomic DNA into DNA nanoballs. Unchained sequencing by ligation is then apply to determine the nucleotide sequence (Drmanac *et al.*, 2010) [7]. This method of DNA sequencing enable large numbers of DNA Nano balls to be sequenced per run with the advantage of cost effectiveness as compare to other high-throughput sequencing platforms (Porreca, *2010)* [3]. with undue advantage that only short sequences of DNA are determined from each DNA nanoball which makes mapping the short reads to a reference genome difficult. This technology has been involved for multiple genome sequencing projects and is scheduled to be used for more purposes (Canard *et al.*, 1994) [4]

Nano pore DNA sequencing

The technique involve DNA passing through the nanopore changes it to ion current. This alteration is dependent on the shape, size and length of the DNA sequence. Each type of the nucleotide obstruct the ion flow through the pore for a different time interval. The method does not involve the use of modified nucleotides and is accomplished in real time. Nanopore sequencing is assign to as "third-generation" or "long-read" sequencing, along with SMRT sequencing. Early this method was based on a technique called 'Exonuclease sequencing', where the readout of electrical signals occurring at nucleotides passing by alpha(α)-hemolysin pores covalently bound with cyclodextrin (Clarke *et al., 2009)* [5]. Two thrust areas of nanopore sequencing in development are solid state

nanopore sequencing, and protein based nanopore sequencing. Protein nanopore sequencing uses membrane protein complexes such as α-hemolysin, MspA (*Mycobacterium smegmatis* Porin A) or CssG, which show great promising ability to distinguish between individual and groups of nucleotides (Dela *et al.*, 2012) [6]. "Fabrication and characterization of solid-state nanopore arrays for high-throughput DNA sequencing". In contrast, solid-state nanopore sequencing involve the use of synthetic materials such as silicon nitride and aluminum oxide which is preferred for its superior mechanical ability and thermal and chemical stability (Pathak *et al.*, 2012) [25]. The fabrication method is necessary for this type of sequencing given that the nanopore array can contain hundreds of pores with diameters smaller than eight nanometers (Dela *et al.*, 2012) [6].

Applications

High-throughput determination of RNA structures

High-throughput sequencing technology is enabling the structures of RNA to be determined at remarkably unparalleled scale, providing deep understanding into the relationship between the structures adopted by RNAs and the functions they conduct in the cell (Strobel *et al.*, 2018) [36].

Towards a genomics-informed, real-time, global pathogen surveillance system

Next-generation sequencing has the capability to assist public health surveillance systems to diagnose the early detection of emerging infectious diseases (Gardy *et al.*, 2018) [10]

Enhancing the accuracy of next-generation sequencing for detecting rare and Subclonal mutations

Despite the exceptional throughput of next-generation sequencing technologies, standard techniques are still confined to distinguishing sequencing errors from genuine low-frequency DNA variants within heterogeneous cellular or molecular populations (Salk *et al.*, 2018) [29].

Harnessing ancient genomes to study the history of human adaptation

Ancient genomes can be utilized in better understanding of the history of human adaptation and evolution through the direct tracing of alteration in genetic variant frequency across different geographical areas and different time periods. (Marciniak & George, 2017) [17].

References

1. Ashley EA, Butte AJ, Wheeler MT, Chen R, Klein TE, Dewey FE. *Et al.* Clinical assessment incorporating a personal genome. The Lancet. 2010; 375(9725):1525-1535.
2. Balasubramanian S. Solexa sequencing: decoding genomes on a population scale. Clinical chemistry. 2015; 61(1):21-24.
3. Brenner S, Johnson M, Bridgham J, Golda G, Lloyd DH, Johnson D. *et al.* Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. Nature biotechnology. 2000; 18(6):630.
4. Canard B, Sarfati S. Novel derivatives usable for the sequencing of nucleic acids. CA Patent. 1994; 2:13.
5. Clarke J, Wu HC, Jayasinghe L, Patel A, Reid S, Bayley H. Continuous base identification for single-molecule nanopore DNA sequencing. Nature nanotechnology. 2009; 4(4):265.
6. dela Torre R, Larkin J, Singer A, Meller A. Fabrication and characterization of solid-state nanopore arrays for high-throughput DNA sequencing. Nanotechnology. 2012; 23(38):385308.
7. Drmanac R, Sparks AB, Callow MJ, Halpern AL, Burns NL, Kermani BG, Human genome sequencing using unchained base reads on self-assembling DNA nanoarrays. Science. 2010; 327(5961):78-81.
8. Eid J, Fehr A, Gray J, Luong K, Lyle J, Otto G, *et al.* Real-time DNA sequencing from single polymerase molecules. Science. 2009; 323(5910):133-138.
9. Fuller CW, Middendorf LR, Benner SA, Church GM, Harris T, Huang X. *et al.* The challenges of sequencing by synthesis. Nature biotechnology. 2009; 27(11):1013.
10. Gardy JL, Loman NJ. Towards a genomics-informed, real-time, global pathogen surveillance system. Nature Reviews Genetics. 2018; 19(1):9.
11. Hart C, Lipson D, Ozsolak F, Raz T, Steinmann K, Thompson J. *et al.* Single-molecule sequencing: sequence methods to enable accurate quantitation. In Methods in enzymology. 2010; 472:407-430.
12. Hickman SE, Kingery ND, Ohsumi TK, Borowsky ML, Wang LC, Means TK. El Khoury J. The microglial sensome revealed by direct RNA sequencing. Nature neuroscience. 2013; 16(12):1896.
13. Huang YF, Chen SC, Chiang YS, Chen TH, Chiu KP, December. Palindromic sequence impedes sequencing-by-ligation mechanism. In BMC systems biology BioMed Central. 2012; 6(2):10.
14. Illumina Sequencer Comparison Table [Internet]. Available. from: http://www.illumina.com/systems/sequencing.html Accessed: 2015-06-16
15. Koren S, Harhay GP, Smith TP, Bono JL, Harhay DM, Mcvey SD, Reducing assembly complexity of microbial genomes with single-molecule sequencing. Genome biology. 2013; 14(9):101.
16. Liu L, Li Y, Li S, Hu N, He Y, Pong R. *et al.* Comparison of next-generation sequencing systems. Bio Med Research International, 2012.
17. Marciniak S, Perry GH, Harnessing ancient genomes to study the history of human adaptation. Nature Reviews Genetics. 2017; 18(11):659.
18. Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Genome sequencing in micro fabricated high-density picolitre reactors. Nature. 2005; 437(7057):376.
19. Maxam AM, Gilbert W. A new method for sequencing DNA. Proceedings of the National Academy of Sciences. 1977; 74(2):560-564.
20. McCoy RC, Taylor RW, Blauwkamp TA, Kelley JL, Kertesz M, Pushkarev D. *et al.* Illumina Tru Seq synthetic long-reads empower de novo assembly and resolve complex, highly-repetitive transposable elements. PloS one. 2014; 9(9):106689.
21. Metzker ML, Sequencing technologies-the next generation. Nature reviews genetics. 2010; 11(1):31.
22. Munroe DJ, Harris TJ, Third-generation sequencing fireworks at Marco Island. Nature biotechnology. 2010; 28(5):426.
23. Nyren, Lundin. Enzymatic method for continuous monitoring of inorganic pyrophosphate synthesis. Anal Biochem. 1985; 151(2):504-509

24. Nyrén P, Pettersson B, Uhlén M. Solid phase DNA minisequencing by an enzymatic luminometric inorganic pyrophosphate detection assay. Analytical biochemistry. 1993; 208(1):171-175.
25. Pathak B, Löfås H, Prasongkit J, Grigoriev A, Ahuja R, Scheicher RH. Double-functionalized nanopore-embedded gold electrodes for rapid DNA sequencing. Applied Physics Letters. 2012; 100(2):023701.
26. Porreca GJ. Genome sequencing on nanoballs. Nature biotechnology. 2010; 28(1):43.
27. Pushkarev D, Neff NF, Quake SR. Single-molecule sequencing of an individual human genome. Nature biotechnology. 2009; 27(9):847.
28. Rizzo JM, Buck MJ, Key principles and clinical applications of" next-generation" DNA sequencing. Cancer Prev Res (Phila). 2012; 5(7):887-900 doi: 10.1158/1940-6207. CAPR-11-0432.[PubMed][Cross Ref].
29. Salk JJ, Schmitt MW, Loeb LA. Enhancing the accuracy of next-generation sequencing for detecting rare and subclonal mutations. Nature Reviews Genetics. 2018; 19(5):269.
30. Sanger F, Coulson AR. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. Journal of molecular biology. 1975; 94(3):441-448.
31. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. Proceedings of the national academy of sciences. 1977; 74(12):5463-5467.
32. Schadt EE, Turner S, Kasarskis A. A window into third-generation sequencing. Human molecular genetics, 2010; 19:227-240.
33. Schuster SC. Next-generation sequencing transforms today's biology. Nat. Methods. 2008; 5(1):16–18.
34. Shendure J, Porreca GJ, Reppas NB, Lin X, Mc Cutcheon JP, Rosenbaum AM, *et al.* Accurate multiplex polony sequencing of an evolved bacterial genome. Science, 2005; 309(5741):1728-1732.
35. Shendure J, Ji H. Next-generation DNA sequencing. Nature biotechnology. 2008; 26(10):1135.
36. Strobel EJ, Angela MY, Lucks JB. High-throughput determination of RNA structures. Nature Reviews Genetics, 2018, 1.
37. Travers KJ, Chin CS, Rank DR, Eid JS, Turner SW, A flexible and efficient template format for circular consensus sequencing and SNP detection. Nucleic acids research. 2010; 38(15):159-e159.
38. Valouev A, Ichikawa J, Tonthat T, Stuart J, Ranade S, Peckham H. *et al.* A high-resolution, nucleosome position map of C. elegans reveals a lack of universal sequence-dictated positioning. Genome research, 2008, 076463.
39. Voelkerding KV, Dames SA, Durtschi JD. Next-generation sequencing: from basic research to diagnostics. Clinical chemistry. 2009; 55(4):641-658.
40. Wetterstrand KA. DNA sequencing costs: data from the NHGRI Genome Sequencing Program (GSP). 2013.