REVIEW ARTICLE



High Throughput Sequencing: An Overview of Sequencing Chemistry

Sheetal Ambardar 1,3,4 · Rikita Gupta 1 · Deepika Trakroo 1 · Rup Lal 2 · Jyoti Vakhlu 1

Received: 16 December 2015/Accepted: 22 June 2016/Published online: 9 July 2016 © Association of Microbiologists of India 2016

Abstract In the present century sequencing is to the DNA science, what gel electrophoresis was to it in the last century. From 1977 to 2016 three generation of the sequencing technologies of various types have been developed. Second and third generation sequencing technologies referred commonly to as next generation sequencing technology, has evolved significantly with increase in sequencing speed, decrease in sequencing cost, since its inception in 2004. GS FLX by 454 Life Sciences/Roche diagnostics, Genome Analyzer, HiSeq, MiSeq and NextSeq by Illumina, Inc., SOLiD by ABI, Ion Torrent by Life Technologies are various type of the sequencing platforms available for second generation sequencing. The platforms available for the third generation sequencing are HelicosTM Genetic Analysis System by SeqLL, LLC, SMRT Sequencing by Pacific Biosciences, Nanopore sequencing by Oxford Nanopore's, Complete Genomics by Beijing Genomics Institute and GnuBIO by BioRad, to name few. The present article is an overview of the principle and the sequencing chemistry of these high throughput sequencing technologies along with brief comparison of various types of sequencing platforms available.

- ☑ Jyoti Vakhlu jyotivakhlu@gmail.com
- Metagenomics Laboratory, School of Biotechnology, University of Jammu, Jammu, J&K, India
- Molecular Biology Laboratory, Department of Zoology, South Campus, University of Delhi, Delhi, India
- Centre for Cellular and Molecular Platform, National Centre for Biological Sciences, TIFR Bangalore, Bangalore, India
- Present Address: Institute of Trans-Disciplinary Health Sciences and Technology, Trans-Disciplinary University, Bangalore 560064, India

Keywords Sequencing platforms · Second generation sequencing · Third generation sequencing · High throughput sequencing · NGS

Introduction

The DNA sequencing methods were developed by Sanger and Coulson and Maxam and Gilbert in 1977 [1, 2]. Initially, both methods were equally popular with researchers but, with invention of thermal cycler, automation and use of nonhazardous chemicals, Sanger's method became more popular. Sanger's sequencing has been evolved into current automated DNA sequencing that is also referred to as the 'First- Generation Sequencing' (FGS), where the terminator ddNTP is tagged with specific fluorescent dyes [3]. Second generation sequencing (SGS), involves massively parallel sequencing of number of templates of same sample in a single run and produces an enormous volume of data economically. SGS has catalyzed the number of breakthroughs, such as advancing scientific knowledge in human disease research to agriculture, microbial ecology to evolutionary science [4–6]. In contrast to the short read length of SGS technologies already available, third generation sequencing (TGS) results in longer read length at low cost. TGS is the ultimate dream of biologists, as it will bring freedom from amplification artifacts and bias [7]. In the sequencing reactions, the nucleotide base assignment is referred to as "Base calling" which is scanned for error probability by Phred software. Phred reads DNA sequence chromatogram files and assigns quality score ("Phred scores") by examining the peaks around each base call. Phred score varies from 4 to 60 with better quality and low error probabilities associated with higher values. Phred/ quality score is used to compare the efficacy of different



sequencing methods [8]. As ample literature is available on the first generation sequencing technology, this article deals with the principle and sequencing chemistry of the second and third generation sequencing technologies, as overviewed in Fig. 1.

Second Generation Sequencing (SGS) Technologies

The 454 Roche GS FLX System, being phased out now, was the first commercially available next generation sequencing platform in 2004 and was, followed by Illumina genome analyzer in 2006, the SOLiD sequencer in 2007 and the Ion Torrent in 2010. Different types of platforms differ primarily in sequencing chemistries that lead to differences in throughput, read length, error rate, genome coverage, cost and run time [5, 6] Two steps involved in all SGS are:

- 1. Template preparation
- 2. Sequencing

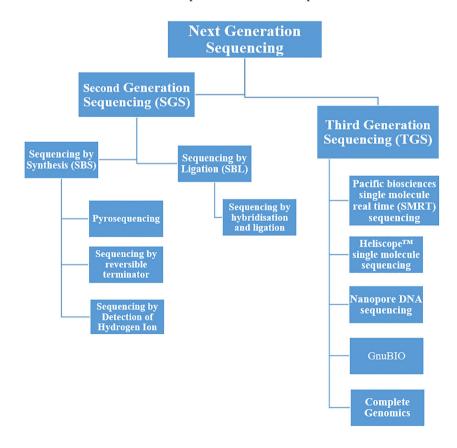
Template Preparation

Template preparation has further three steps:

A. Source nucleic acid extraction

Fig. 1 Overview of the second and third generation sequencing technologies

- B. Library preparation
- C. Template amplification
- (A) Source nucleic acid extraction Nucleic acid extraction protocols are not universal but depend upon the sample source and type of study to be conducted [5, 9].
- (B) Library preparation DNA library prepared for sequencing should not be confused with genomic/cDNA libraries. DNA Library preparation here involves fragmentation of isolated DNA into smaller, random, overlapping fragments followed by the end polishing of fragmented DNA and adapter ligation. The isolated DNA is fragmented in the range from 150 to 800 bp depending on the platform to be used and is done by one of the three following methods i.e. physical/mechanical methods (i.e., Nebulization and ultrasonication) or enzymatic methods (i.e., non-specific endonuclease cocktails) and transposase (tagmentation reactions) [10–13]. Library preparation from RNA is done by capturing mRNA, random priming and complementary DNA (cDNA) synthesis followed by the end polishing and adapter ligation [14–16]. Since the average size of the RNA is smaller, so fragmentation is usually not required in RNA/transcriptome sequencing.
- (C) Template amplification DNA fragments selected after library preparation or specific amplification are subjected to clonal amplification. Clonal amplification involves solid





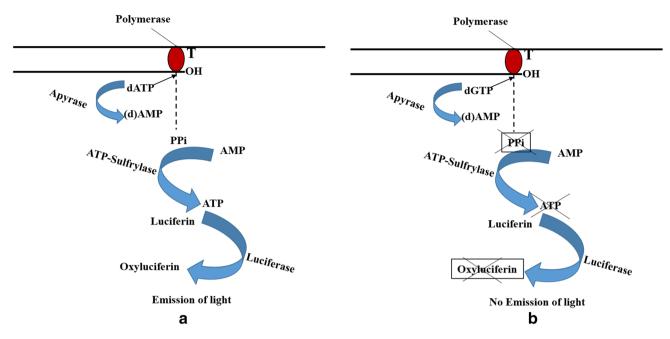


Fig. 2 Diagrammatic representation of enzymatic reactions involved in Pyrosequencing a during complementary nucleotide incorporation and b when nucleotide is not incorporated

phase amplification of DNA fragments and helps in development of strong detectable signal during sequencing. Single DNA fragment to be sequenced is either bound to beads, ion surfaces or flow cell. Depending upon the sequencing platform, emulsion PCR (emPCR) or bridge PCR is used to amplify the anchored DNA fragments into millions of spatially separated template fragments [17, 18].

Sequencing

The difference in various SGS technologies lies in the principle of sequencing each base. There are two basic sequencing approaches developed so far:

- 1. Sequencing, by synthesis (SBS)
- 2. Sequencing, by hybridization and ligation (SBL)

Sequencing by Synthesis (SBS)

The three main type of sequencing chemistry followed in SBS are as under:

- A. Pyrosequencing
- B. Sequencing by reversible termination
- C. Sequencing by detection of hydrogen ions
- (A) Pyrosequencing After preparation of the template by emPCR [17], the clonally amplified beads entrapped in the wells of picotitre plate, are ready for pyrosequencing. In Pyrosequencing, pyrophosphate (PPi)

released during the DNA polymerization reaction is detected and used as an indicator of specific base incorporation. It involves detection of the base incorporated in the daughter DNA by DNA polymerase, with feeding and removing, all the four bases sequentially and detecting the pyrophosphate released by cascade of enzymes that emit light [7, 19].

Single stranded template DNA attached to bead, hybridizes with sequencing primer and is incubated with DNA polymerase, ATP sulfurylase, Luciferase, Apyrase, Adenosine 5' phosphate (APS), luciferin and dNTPs in the well of the plate. dATP is replaced by dATPαS, in which one of the oxygen in S-position of the α -phosphate, is replaced by Sulfur, as former is substrate for luciferase. On the incorporation complementary dNTPs by DNA polymerase, pyrophosphate (PPi) is released and is converted into ATP by ATP sulphurylase using adenosine 5'phosphosulfate. In the presence of ATP, luciferase converts luciferin into oxyluciferin that generates visible light (Fig. 2). The light produced by luciferase reaction is detected and measured by an avalanche photodiode, photomultiplier tube, or with a charge-coupled device camera (with or without a microchannel plate).

The limitation of pyrosequencing is, inaccurate homopolymer sequencing, as addition of more than five identical nucleotides cannot be detected efficiently [19]. Additionally, the cost in comparison to other high throughput sequencing technologies is comparatively high. Although a pioneer in second generation sequencing technologies, pyrosequencing has already been pushed out



of the race and this indicates the rate at which sequencing technologies are evolving.

(B) Sequencing by reversible termination With Illumina sequencing platform, DNA fragments of the libraries are subjected to clonal amplification by bridge PCR [18] followed by sequencing by reversible termination using reversible terminator (RT) nucleotide. The RT nucleotide is protected at 3'-OH groups (2-cyanoethyl) and is fluorescently labeled [20]. On addition of mixture of RT nucleotides to the flow cell, the DNA polymerase incorporates modified nucleotides into the DNA strand being synthesized. Each cycle in sequencing by reversible termination consists of three steps, i) incorporation of the complementary RT nucleotide by mutant DNA polymerase to the DNA strand attached with flowcell followed by, ii) detection of the different fluorescence signal for the four bases and iii) finally restoration of free 3'OH group by cleaving the terminating moiety and reporter molecule. Repetition of this cycle leads to sequencing of the DNA template. During sequencing, RT nucleotide is incorporated and imaged followed by removal of fluorophores and terminating bases are activated by de-protection at the 3'OH group, allowing another round of nucleotide incorporation (Fig. 3). Natural competition between all the four nucleotides present during each sequencing cycle reduces the inherent bias as compared to pyrosequencing where only one type of nucleotide is made available at a time for pairing [20, 21]. Homopolymers sequencing error (as in case of pyrosequencing) is overcome by this technique due to incorporation of single base at a time, as for addition of another base terminator needs to be removed first. Fourchannel sequencing system is used in Illumina HiSeq and MiSeq wherein each base is detected by individual image. Illumina has come up with the latest NextSeq 500 that has only two-channel SBS technology that requires only two images to determine all four base calls. This reduces the image capturing time, number of cycle, cost of sequencing and time required for data processing, while delivering high quality and accuracy [22].

Only Illumina NGS platforms are capable of paired-end sequencing as the clonal amplification here is done by bridge PCR [18]. Paired –end sequencing allows users to sequence the DNA fragment from both the ends resulting in high coverage, high numbers of reads and more data as compared to single end sequencing systems. Paired-end sequencing generates high-quality sequence data due to increased probability of alignment to a reference. It facilitates detection of genomic rearrangements (insertions, deletions, and inversions), repetitive sequence elements, gene fusions and novel transcripts. In addition it also provides superior alignment across DNA regions containing repetitive sequences, and produce longer contigs for de novo sequencing by filling gaps in the consensus sequence. In transcriptome sequencing (RNA-Seq) it enables discovery applications such as detecting gene fusions in cancer and characterizing novel splice isoforms [23].

Another technology upgradation offered by Illumina in 2015 is patterned flowcells in HiSeq X Ten, HiSeq 3000/ HiSeq 4000 systems. This provides exceptional level of throughput due to billions of nanowells at fixed locations as compared to normal flowcell. The structured organization of the patterned flow cell provides even cluster spacing and uniform feature size that will allow binding of single DNA template within a single well for cluster formation. This results in high well occupancy and maximum data output, many folds than original HiSeq and MiSeq [24].

Illumina in partnership with 10× Genomics Gem/Code technology has been able to develop technology that generates read length up to 100 kb, comparatively more than previous illumina sequencing platforms to enable long read

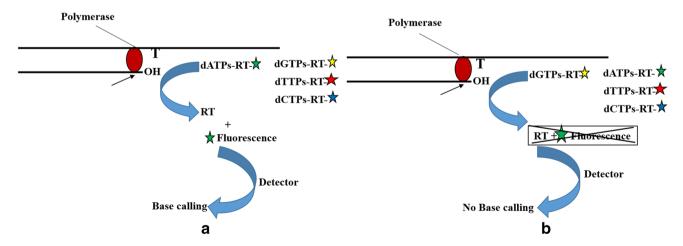


Fig. 3 Diagrammatic representation of Sequencing by reversible termination \mathbf{a} during complementary nucleotide incorporation and \mathbf{b} when nucleotide is not incorporated



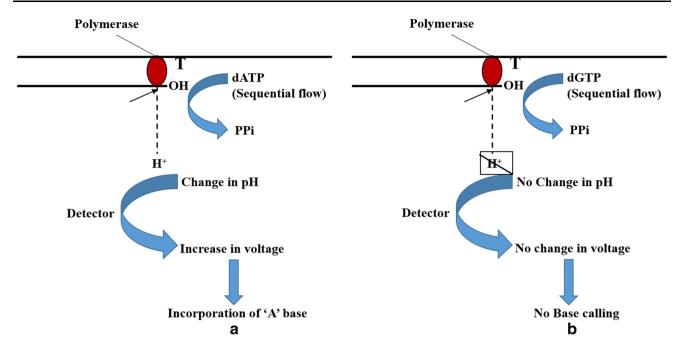


Fig. 4 Diagrammatic representation of pH change involved in Sequencing by detection of hydrogen ions a during complementary nucleotide incorporation and b when nucleotide is not incorporated

applications. Illumina's 10× Genomics technology/Gem-Code platforms uses microfluidic system to partition DNA samples (100 Kb or more) into gems to increase the read length in comparison to original HiSeq, NextSeq and MiSeq platforms. The microfluidic system includes highthroughput, droplet-based reagent delivery system that uses hydrogel beads (gel beads) to deliver barcoded oligonucleotides [25]. Cartridge reservoirs are loaded with gel beads, sample & reagent mixture, oil-surfactant solution and it can process 8 samples at a time. The cartridge reservoirs delivers the reagents via a network of microfluidic channels, wherein oil based droplets are formed having DNA fragment of (100 Kb), reagents and the gel bead having millions of copies of the same barcoded oligonucleotide. The barcode gets incorporated in the longer DNA fragments using GemCode instrument and reagents from 10× Genomics, followed by library preparation and sequencing on the Illumina HiSeq sequencers. GemCode technology can be utilized by HiSeq XTM Ten, HiSeq[®], NextSeq[®] and Miseq[®] sequencer. It offers 140 % increase in speed and 50 % more clusters per run compared to HiSeq 2500 and HiSeq 2000 respectively. GemCode software is used after sequencing for assembly of long reads (100 Kb or more) from shorter reads [25]. The short reads having same barcodes are assembled to make larger fragments thereby resulting in long sequence read.

(C) Sequencing by detection of hydrogen ions This is based on detection of hydrogen ion liberated on incorporation of each nucleotide and is not dependent upon altered bases, enzymes or optical detection [26]. Template DNA obtained after the library preparation and clonal amplification is bound to the proprietary Ion Sphere particles in the microwell in such a way that each microwell contains a single Ion sphere particles. A single type of dNTP is added to the microwell at a time and its incorporation is detected by libration of hydrogen ion that triggers an ISFET (Ion Sensitive Field Effect Transistor) ion sensor. The change in pH is detected by a sensing layer of microwell which translates the chemical signal into digital signal, measured within seconds (Fig. 4). In comparison to other sequencing methods, where detection is indirect using laser scanners or CCD cameras, the detection here is independent of these devices and hence direct [27]. This sequencing is also referred to as Ion Torrent sequencing, pH mediated sequencing, Silicon sequencing or Semiconductor sequencing. This technique is rapid as sequencing is done in real time with read length of 200 nucleotides and operates at low cost since it does not incorporate modified bases and utilizes a single polymerase enzyme [27, 28]. However, similar to Roche 454 in this sequencing chemistry, homopolymer sequencing is error prone as repeats will be incorporated in one cycle which leads to a proportionally higher electronic signal due to corresponding number of released hydrogen ion. Life Technologies have come up with Hi-Q polymerase that has improved sequencing accuracy, decreased indel error rates and reduces GC-bias. Ion Hi-Q sequencing is useful for whole-genome and transcriptome sequencing. [29].



Sequencing by Hybridization and Ligation (SBL)

Sequencing by ligation is basis of support oligonucleotide ligation detection (SOLiD) sequencing platform by Applied Biosystems. SBL depends upon the specificity of the DNA ligase for base-pair mismatch instead of DNA polymerase in case of SBS. Template preparation includes fragmentation of template and its attachment with known adapter sequence. The adapters attached to the DNA fragment to be sequenced is further attached to beads and clonally amplified by emPCR [17]. These clonally amplified DNA attached to beads are deposited on glass plate where the beads get covalently bound [28, 30].

Sequencing chemistry includes the hybridisation and ligation of various one/two -base-encoded probes to the template. The probe is eight or nine bases long and consists of one or two base followed by three degenerate bases and three universal bases which are attached to fluorescent label. A short primer is added along with the mixed pool of fluorescent oligonucleotide probe that hybridizes/anneal to the target DNA having complementary sequence. The probes are ligated with the primers using a DNA ligase and are detected by fluorescence imaging and non-ligated probes are washed away. The oligonucleotide probes have cleavable linkages attached to fluorescent label which can be cleaved after detection thereby preparing the system for another round of ligation (Fig. 5). This cycle is repeated several times to sequence the complete target DNA. At the end of one round, the sequences of bases will be known at only some positions other than the degenerate bases. To sequence the skipped positions, another round of

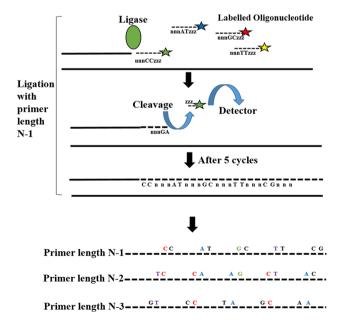


Fig. 5 Diagrammatic representation of enzymatic reactions involved in Sequencing by hybridization and ligation

sequencing occurs with a primer that has one or more bases shorter than the previous primer. Thus, the sequence of complete target DNA can be obtained using anchors of various length [28, 30].

The read length of SOLiD chemistry is a major limitation of this chemistry but has been improved from 35 bp to 85 bp till now. Short read lengths lead to inaccuracy in read assembly as it is difficult to assemble with any pipeline and additionally more time is required for sequencing [30]. However modified platform for sequencing by hybridization has been developed that will be dealt in third generation sequencing technologies (section on Complete Genomic technology).

Third Generation Sequencing Approaches

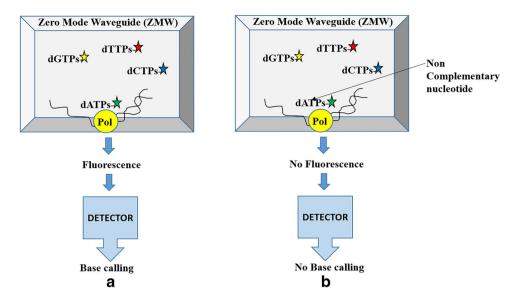
Second generation sequencing platforms in general suffer from two limitations i) the short read length that need to be assembled with the help of various bioinformatic tools/ pipelines into original length template and ii) PCR bias introduced by clonal amplification, for detection of base incorporation signal. The "third generation of high throughput NGS technology" or single molecule real time (SMRT) sequencing has been developed as a remedy to these limitations [31]. Instead of sequencing clonally amplified template, single DNA template is sequenced and this has also led to minimal use of biochemicals leading to miniaturization of whole process to nanoscale. Third generation sequencing till date comprises of five sequencers wherein only Helicos' Genetic Analysis System, Pacific Biosciences and Oxford Nanopore's nanopore sequencing are single molecule real time technology platforms. [32]. The other two TGS platform are Complete Genomics by Beijing Genomics Institute and GnuBio by BioRad. Complete Genomics is based on hybridization and ligation whereas GnuBio is based on microfluidics which combines PCR amplicons from a genomic sample with nanodroplets containing one of ~ 5000 different hexameric sequencing primers, reducing the number of steps into single high throughput library preparation step.

Pacific Biosciences Single Molecule Real Time (SMRT) Sequencing

Pacific Biosciences in 2010, commercialized an approach combining nanotechnology with molecular biology to sequence single molecule known as single molecule real time (SMRT) sequencing. In this technology, instead of immobilizing DNA strands, the high fidelity φ 29 derived DNA polymerase along with a single strand DNA template is immobilized at the bottom of zero mode waveguides (ZMW). Zero-mode waveguide (ZMW) is a



Fig. 6 Diagrammatic representation of single molecule real time (SMRT) sequencing in Zero-mode waveguide a during complementary nucleotide incorporation and b when nucleotide is not incorporated



nanophotonic confinement structure that consists of a circular hole in an aluminum cladding film deposited on a clear silica substrate. The ZMW holes are ~70 nm in diameter and ~ 100 nm in depth. Due to the behavior of light when it travels through a small aperture, the optical field decays exponentially inside the chamber. The observation volume within an illuminated ZMW is ~20 zeptoliters (20×10^{-21} L). Within this volume, the activity of DNA polymerase incorporating a single nucleotide can be readily detected. During sequencing, φ29 DNA polymerase incorporates the dye labeled nucleotides and the signal is detected by instrument optics which continuously monitors enzyme's active site in each zero mode waveguides (ZMW) (Fig. 6). The Pacific Biosciences sequencer can generate the read length of ~40 Kb but at 85 % accuracy lower than Second Generation sequencers. The error rate of 15 % is a consequence of detection of nucleotide dwelling into the active site for long enough even if it is not subsequently incorporated into the DNA strand [33].

HelicosTM Single Molecule Sequencing

The HelicosTM single molecule sequencing is done on HeliScope Genetic Analysis System platform marketed by SeqLL, LLC. The library preparation of Helicos sequencing is easier than others sequencing technology as it does not require ligation and amplification for library preparation. Instead, the DNA is sheared, tailed with poly-A, which is further blocked at 3′ OH end using terminal transferase and a dideoxynucleotide. The poly-A fragments are hybridized to the flow-cell surface containing oligo-dT for initiating sequencing-by-synthesis. HeliScope sequencer uses fluorescent tagged nucleotides for sequencing the DNA fragments that are attached to flowcell through poly T tails (Fig. 7). This is the first

commercial TGS to use the principle of single molecule fluorescent sequencing. It also allows sequencing and quantitation of RNA molecules (without converting them into cDNA) involving direct RNA hybridization to the flow cell. This sequencing technology is in its infancy due to small read length (24-70bases) and low data output (20 Gb) [34, 35].

Nanopore DNA Sequencing

Yet another single molecule TGS developed by Oxford Nanopore technologies reads incorporating nucleotides using nanopore technology. As the DNA sequence passes through a nanopore having an internal diameter of 1 nm, the electrical conductance of the pore is altered and signal is detected (Fig. 8). This technology involves the use of protein nanopores embedded in the polymer membrane [35]. The sequencing does not need any intervening PCR amplification or a chemical labeling step. In addition there is no need for sample preparation as cell lysate can be directly sequenced [15]. The Company introduced its pocket-sized sensing device MinION in 2014 and has been commercially made available in May 2015. MinION nanopore sequencer's performance has been optimized using M13 genomic DNA and 99 % of reads were mapped to reference genome [36, 37]. This sequencer has 512-2000 nanopores with each nanopore having the sequencing speed of 120-1000 bases per minute. The sequencer is like a USB stick and can be used only one time. This technology will make all the sequencing machines used till date redundant and, reduce the cost & effort of sequencing tremendously [36]. Instead of collecting samples and sequencing them in lab, the sequencers will be taken to field and sequencing will be done in field directly [15, 37, 38]. However with the



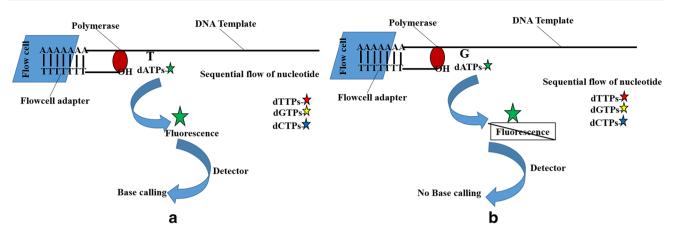


Fig. 7 Diagrammatic representation of Helicos single molecule sequencing a during complementary nucleotide incorporation and b when nucleotide is not incorporated

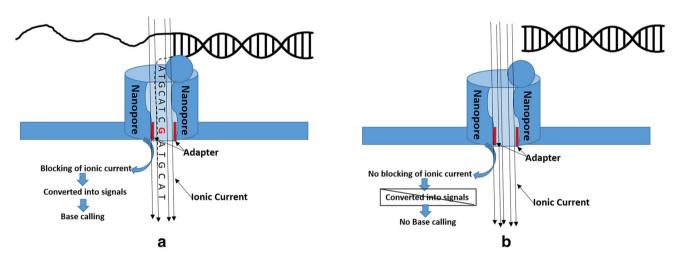


Fig. 8 Diagrammatic representation of a nanopore sequencing during base calling and b no base case calling

current error rate of 38.2 % it needs more improvement on this front [39].

Complete Genomics Technology

Complete Genomic technology is sequencing platform based on hybridization and ligation developed by Complete Genomics. Library construction involves insertion of four-adaptors of ~ 70 bases within genomic DNA at regular intervals which is known as Combinatorial Probe-Anchor Ligation (cPAL TM) technology that results in increased read length. The read length can be further increased by additional numbers of adaptors. The clonal amplification results in DNA clusters termed as DNA nano-balls (DNBs) that occurs in a single reaction. Clonally amplified DNA on nano-balls is further sequenced. cPAL technology has overcome three limitations of SOLiD namely, short read length, error in the accuracy of repetitive base sequencing and sequence analysis [40].

GnuBIO

GnuBIO, a Bio-Rad life sciences company has developed a droplet-based DNA sequencing platform. The sequencing platform utilizes microfluidic and emulsion technology to perform the complex reactions in droplets, thereby reducing the number of steps into single high throughput library preparation step. The microfluidic workflow combines PCR amplicons from a genomic sample with nanodroplets containing one of ~ 5000 different hexameric sequencing primers, each associated with a specific dye barcode. By using an algorithm to determine which hexamers do or do not hybridize to a given amplicon, the system's on-board computer can map sequence and structural irregularities with remarkable accuracy [41]. The science behind this GnuBIO sequencing technology utilizes droplet microfluidics to perform the biochemical reactions for sequencing inside of tiny picoliter -sized aqueous drops. Each droplet



Table 1 Comparison of various specifications of different NGS technologies

Sequencing platforms	Error rate (%)	Read length (nts)	No. of million reads	Yield Gb/run	Reference
Pyrosequencing (454 Roche)	1	500	1	0.5	[19, 43]
Sequencing by reverse terminator (Illumina Hiseq 2500)	0.26	2×100	8000 PE	720-800	[20-24]
Sequencing by reverse terminator (Illumina Hiseq 2500 Rapid Run)	0.26	2×250	1200 PE	150-180	
Sequencing by reverse terminator (Illumina NextSeq)	0.8	2×150	800 PE	100-120	
Sequencing by reverse terminator (Illumina Miseq)	0.8	2×300	44-50 PE	13.2-15	
Sequencing by reverse terminator (Illumina MiniSeq)	0.8	2×150	50	6.5 - 7.5	
Sequencing by detection of hydrogen ion (Ion Torrent)	1.78	200	80	10	[26-29]
Sequencing by ligation SOLiD	0.01	35	1400	155	[28, 30]
Single molecule real time sequencing (Pacific Biosciences)	13	40,000	0.1	0.1	[32, 33]
Oxford nanopore technologies (minION)	38.2	2000	0.03	1	[36–38]
Heliscope TM single molecule sequencing	0.5	25	12-20	35	[34, 35]
Complete genomic technology	1	2×35	NA	NA	[40]

works as a unique reaction vesicle providing a streamlined workflow that decreases reagent cost, facilitates kinetics of reactions, and provides a uniquely scalable desktop sequencing platform. A single GnuBIO instrument encompasses all of the steps required for DNA sequencing into a single platform including target selection/enrichment, DNA amplification, DNA sequencing and analysis. While other sequencing technologies require separate workflows for target selection, DNA amplification, DNA sequencing and analysis [41, 42].

The race to increase the read length, reduce cost & steps involved in sequencing is far from over. Almost every few month new technology or sequencing platform is being introduced and by the time this article is published, there will be few more. Recently, Illumina has launched Tru-Sight HLA typing kit and NeoPrep System (2015) wherein the NeoPrep is only compatible with Illumina technology. TruSight HLA helps in allele specific sequencing and Neoprep is the first commercially available product that performs library prep, quantification and normalization. Complete Genomics has also introduced a new system (2015), i.e., the RevolocityTM system that has inbuilt capacity to perform DNA extraction, library preparation, sequencing and data analysis. This is first sequencing solution for large-scale, high-quality genomes sequencing [43-45].

Limitations of Next Generation Sequencing

Comparison between NGS and FGS leads to supremacy of NGS in terms of cost, labor and speed but as far as error rate and read length is concerned, FGS based on Sanger sequencing is still a gold standard [43]. Different NGS technologies produce different kinds of errors such as

substitution, Indel (insertion/deletion), AT bias, deletions and GC deletions in case of data obtained from capillary/ Illumina, 454 sequencing/Ion Torrent, SOLiD, oxford nanopore and Pacific Biosciences respectively (Table 1). The final error rates range between 0.1 % and 1 % for Capillary, 0.1 % for SOLiD, 0.4 % for the Illumina platforms, 1 % for Pyrosequencing, 1.78 % for Ion Torrent, 15 % for Pacific Biosciences and 38.2 % for minION oxford nanopore technologies [43]. The generation time in development of sequencing technologies is very short and at the same time, the evolution is taking place at a fast rate. Third generation sequencing platforms with single molecule real time sequencing are available and work on error reduction is in progress [45, 46]. It seems that soon the major drawback of short read length and high error rate in third generation sequencing platforms will be history.

Conclusion

The first generation sequencing initially given by Sanger in 1977 has evolved into second generation sequencing (NGS) in 2004 over the span of ~30 years and in another 10 years we have third generation of sequencing platforms developed. NGS has brought a breakthrough development for both basic and applied research in molecular studies. Further developments are in progress for making personalized genomics and community analysis a commonality. NGS/SGS technologies available have advantages and limitations of their own but the low cost and huge data generation is common to all of them. The continuously developing and evolving third generation technologies will soon push all the second generation technologies out of race and the first causality has been pyrosequencing. However scientists working in the field of life sciences are



spoiled for choice as far as sequencing technologies are concerned.

Acknowledgments Authors are thankful to DBT (BT/PR5534/PBD/16/1006/2012), UGC (42-168/2013(SR)) and ICAR-NBAIM (NBAIM/AMAAS/2014-15/81) for funding of various projects. SA and RG are thankful to UGC-CSIR for NET Fellowship.

References

- Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. PNAS 74:5463–5467. doi:10.1073/ pnas.74.12.5463
- Maxam AM, Gilbert W (1977) A new method for sequencing DNA. PNAS 74:560–564. doi:10.1073/pnas.74.2.560
- Quail MA, Smith M, Coupland P, Otto TD, Harris SR, Connor TR, Bertoni A, Swerdlow HP, Gu Y (2012) A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and IlluminaMiSeq sequencers. BMC Genom 13:13. doi:10.1186/1471-2164-13-341
- Metzker ML (2005) Emerging technologies in DNA sequencing. Genome Res 15:1767–1776. doi:10.1101/gr.3770505
- Metzker ML (2009) Sequencing technologies: the next generation. Nat Rev Genet 11:31–46. doi:10.1038/nrg2626
- Scholz MB, Lo CC, Chain PSG (2012) Next generation sequencing and bioinformatic bottlenecks: the current state of metagenomic data analysis. Curr Opin Biotechnol 23:9–15. doi:10.1016/j.copbio.2011.11.013
- Marzorati M, Maignien L, Verhelst A, Luta G, Sinnott R, Kerckhof FM, Possemiers S (2013) Barcoded pyrosequencing analysis of the microbial community in a simulator of the human gastrointestinal tract showed a colon region-specific microbiota modulation for two plant-derived polysaccharide blends. Antonie Van Leeuwenhoek 103:409–420. doi:10.1007/s10482-012-9821-0
- Patel RK, Jain M (2012) NGS QC toolkit: a toolkit for quality control of next generation sequencing data. PLoS ONE 7:e30619. doi:10.1371/journal.pone.0030619
- Fierer N, Lauber CL, Ramirez KS, Zaneveld J, Bradford MA, Knight R (2012) Comparative metagenomic, phylogenetic and physiological analyses of soil microbial communities across nitrogen gradients. ISME J 6:1007–1017. doi:10.1038/ismej. 2011.159
- Caruccio N (2011) Preparation of next-generation sequencing libraries using NexteraTM technology: simultaneous DNA fragmentation and adaptor tagging by in vitro transposition. Method Mol Biol 733:241–255. doi:10.1007/978-1-61779-089-8
- Knierim E, Lucke B, Schwarz JM, Schuelke M, Seelow D (2011) Systematic comparison of three methods for fragmentation of long-range PCR products for next generation sequencing. PLoS ONE 6:e28240. doi:10.1371/journal.pone.0028240
- Parkinson NJ, Maslau S, Ferneyhough B, Zhang G, Gregory L, Buck D, Ragoussis J, Ponting CP, Fischer MD (2012) Preparation of high-quality next-generation sequencing libraries from picogram quantities of target DNA. Genome Res 22:125–133. doi:10. 1101/gr.124016.111
- Hingamp P, Grimsley N, Acinas SG, Clerissi C, Subirana L et al (2013) Exploring nucleo-cytoplasmic large DNA viruses inTara Oceans microbial metagenomes. ISME Journal 7:1678–1695. doi:10.1038/ismej.2013.59
- Frias-Lopez J, Shi Y, Tyson GW, Coleman ML, Schuster SC, Chisholm SW et al (2008) Microbial community gene expression in ocean surface waters. Proc Natl Acad Sci USA 105:3805– 3810. doi:10.1073/pnas.0708897105

- Shi Y, Tyson GW, Eppley JM, DeLong EF (2011) Integrated metatranscriptomic and metagenomic analyses of stratified microbial assemblages in the open ocean. ISME J 5:999–1013. doi:10.1038/ismej.2010.189
- Lesniewski RA, Jain S, Anantharaman K, Schloss PD, Dick GJ (2012) The metatranscriptome of a deep-sea hydrothermal plume is dominated by water column methanotrophs and lithotrophs. ISME J 6:2257–2268. doi:10.1038/ismej.2012.63
- Shao K, Ding W, Wang F, Li H, Ma D, Wang H (2011) Emulsion PCR: a High Efficient Way of PCR Amplification of Random DNA Libraries in Aptamer Selection. PLoS ONE 6:e24910. doi:10.1371/journal.pone.0024910
- Kawashima Eric H, Laurent Farinelli; Pascal Mayer (2005-05-12). "Patent: Method of nucleic acid amplification". Retrieved 2012-12-22
- Fakruddin M, Chowdhury A, Hossain M, Mannan KSB, Mazumdar RM (2012) Pyrosequencing-principles and applications. Life 50:65
- Berglund EC, Kiialainen A, Syvänen AC (2011) Next-generation sequencing technologies and applications for human genetic history and forensics. Investig Genet 2:23. doi:10.1186/2041-2223-2-23
- Chen Y, Sonnaert M, Roberts SJ, Luyten FP, Schrooten J (2012)
 Tissue engineering part C. Methods 18:444–452. doi:10.1089/ten. tec.2011.0304
- Reuter JA, Spacek DV, Snyder MP (2015) High-Throughput Sequencing Technologies. Mol Cell 58:586–597. doi:10.1016/j. molcel.2015.05.004
- http://www.illumina.com/technology/next-generation-sequencing/ paired-end-sequencing_assay.html
- Dijk EL, Auger H, Jaszczyszyn Y, Thermes C (2014) Ten years of next-generation sequencing technology. Trends Genetics 30:418–426. doi:10.1016/j.tig.2014.07.001
- Eisenstein M (2015) Startups use short-read data to expand longread sequencing market. Nat Biotechnol 33:433–435. doi:10. 1038/nbt0515-433
- Eid J, Fehr A, Gray J, Luong K, Lyle J, Otto G, Peluso P, Rank D, Baybayan P, Bettman B et al (2009) Real-time DNA sequencing from single polymerase molecules. Science 323:133–138. doi:10. 1126/science.1162986
- Mascher M, Amand PS, Stein N, Poland J (2013) Application of genotyping-by-sequencing on semiconductor sequencing platforms: a comparison of genetic and reference-based marker ordering in barley. PLoS ONE 8:e76925. doi:10.1371/journal.pone.0076925
- Meldrum C, Doyle MR, Tothill RW (2011) Next-Generation Sequencing for Cancer Diagnostics: a Practical Perspective. Clin Biochem Rev 32:177–195. PMCID: PMC3219767
- 29. Veras AAO, de Sál PHCG, Pinheiro KC, das Graças DA, Baraúna RA, Schneider MPC, Azevedo V, Ramos RTJ, Silva A (2014) Efficiency of *Corynebacterium pseudotuberculosis* 31 Genome Assembly with the Hi-Q Enzyme on an Ion Torrent PGM Sequencing Platform. J Proteomics Bioinform 7:12. doi:10.4172/jpb.1000342
- Liu L, Li Y, Li S et al (2012) Comparison of next-generation sequencing systems. J Biomed Biotechnol vol. 2012, Article ID 251364, 11 pages, 2012. doi:10.1155/2012/251364
- Schadt EE, Turner S, Kasarskis Andrew (2010) A window into third-generation sequencing. Hum Mol Genet 2010:R227–R240. doi:10.1093/hmg/ddq416
- 32. Rusk N (2009) Cheap third-generation sequencing. Nat Methods 6:244. doi:10.1038/nmeth0409-244a
- Mardis ER (2013) Next-generation sequencing platforms. Annu Rev Anal Chem 6:287–303. doi:10.1146/annurev-anchem-062012-092628
- 34. Harris TD, Buzby PR, Babcock H et al (2008) Single-molecule DNA sequencing of a viral genome. Science 320:106–109. doi:10.1126/science.1150427



- Hart C, Lipson D, Ozsolak F, Raz T, Steinmann K, Thompson J, Milos PM (2010) Single molecule sequencing:sequence method to enable accurate quantitation. Methods Enzymol 472:407–430. doi:10.1016/S0076-6879(10)72002-4
- 36. Hayden EC (2012) Nanopore genome sequencer makes its debut. Nature. doi:10.1038/nature.2012.10051
- 37. Jain M, Fiddes IT, Miga KH, Olsen HE, Paten B, Akesen M (2015) Improved data analysis for the MinION nanopore sequencer. Nat Methods 12:351–356. doi:10.1038/nmeth.3290
- Miles BN, Ivanov AP, Wilson KA, Doğan F, Japrung D, Edel JB (2013) Single molecule sensing with solid-state nanopores: novel materials, methods, and applications. Chem Soc Rev 42:15–28. doi:10.1039/c2cs35286a
- Laver T, Harrison J, O'Neill PA, Moore K, Farbos A, Paszkiewicz K, Studholme DJ (2015) Assessing the performance of the Oxford Nanopore Technologies MinION. Biomol Detect Quantif 3:1–8. doi:10.1016/j.bdq.2015.02.001
- http://nextgenseek.com/2012/07/complete-genomics-new-technol ogy-to-produce-accurate-sequencing-with-small-amount-of-dna/
- Basu A, Macosko E, Shalek A, McCarroll S, Regev A, and Weitz D (2014) Single-cell genomics using droplet-based microfuidics.

- Bull Am Phys Soc: APS March Meeting 2014 59:3-7, Denver, Colorado
- 42. Erlich Y (2015) A vision for ubiquitous sequencing. BioRxiv. doi:10.1101/019018
- 43. Rieber N, Zapatka M, Lasitschka B, Jones D, Northcott P, Hutter B et al (2013) Coverage Bias and Sensitivity of Variant Calling for Four Whole-genome Sequencing Technologies. PLoS ONE 8:e66621. doi:10.1371/journal.pone.0066621
- 44. Patnaik BB, Park SY, Kang SW, Hwang H-J, Wang TH, Park EB et al (2016) Transcriptome Profile of the Asian Giant Hornet (Vespa mandarinia) Using Illumina HiSeq 4000 Sequencing: De Novo Assembly, Functional Annotation, and Discovery of SSR Markers. Int J Genomics, 2016:4169587. http://doi.org/10.1155/2016/4169587
- Rosenberg AZ, Armani MD, Fetsch PA, Xi L, Pham TT, Raffeld M et al (2016). High-Throughput Microdissection for Next-Generation Sequencing. PLoS ONE 11:e0151775. http://doi.org/ 10.1371/journal.pone.0151775
- Yu P, Lin W (2016). Single-cell transcriptome study as big data. Genomics Proteomics Bioinf 14:21–30. http://doi.org/10.1016/j. gpb.2016.01.005

