

Generations of Sequencing Technologies: From First to Next Generation

Kchouk Mehdi, Jean-Francois Gibrat, Mourad Elloumi

▶ **To cite this version:**

Kchouk Mehdi, Jean-Francois Gibrat, Mourad Elloumi. Generations of Sequencing Technologies: From First to Next Generation. Electromagnetic Biology and Medicine, 2017, 9 (3), 8 p. 10.4172/0974-8369.1000395 . hal-02621960

HAL Id: hal-02621960
https://hal.inrae.fr/hal-02621960

Submitted on 26 May 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Biology and Medicine
ISSN: 0974-8369

Kchouk et al., Biol Med (Aligarh) 2017, 9:3
DOI: 10.4172/0974-8369.1000395

Review Article OMICS International

Generations of Sequencing Technologies: From First to Next Generation

Mehdi Kchouk[1,3*], **Jean-François Gibrat**[2] **and Mourad Elloumi**[3]

[1]*Faculty of Sciences of Tunis (FST), Tunis El-Manar University, Tunisia*

[2]*Research Unit Applied Mathematics and Computer Science, from Genomes to the Environment' (MaIAGE), Jouy en Josas, France*

[3]*Laboratory of Technologies of Information and Communication and Electrical Engineering, National Superior School of Engineers of Tunis, University of Tunis, Tunisia*

[*]**Corresponding author:** Mehdi Kchouk, Faculty of Sciences of Tunis (FST), Tunis El-Manar University, Tunisia, Tel: +216 71 872 253; E-mail: mehdi.kchouk@gmail.com

Received date: January 31, 2017; **Accepted date:** February 27, 2017; **Published date:** March 06, 2017

Copyright: © 2017 Kchouk M, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution and reproduction in any medium, provided the original author and source are credited.

Abstract

DNA sequencing process utilizes biochemical methods in order to determine the correct order of nucleotide bases in a DNA macromolecule using sequencing machines. Ten years ago, sequencing was based on a single type of sequencing that is Sanger sequencing. In 2005, Next Generation Sequencing Technologies emerged and changed the view of the analysis and understanding of living beings. Over the last decade, considerable progress has been made on new sequencing machines. In this paper, we present a non-exhaustive overview of the sequencing technologies by beginning with the first methods history used by the commonly used NGS platforms until today. Our goal is to provide beginners in the field as well as to the amateurs of science a simple and understandable description of NGS technologies in order to provide them with basic knowledge as an initiation into this field in full ardor.

Keywords: Next generation sequencing technologies; DNA; Sequencing; Long reads; Short reads

Introduction

The discovery of the double helix structure composed of four Deoxyribonucleic Acid (DNA) bases {A, T, C, G} by Watson JD et al. in 1953 [1] has led to the decoding of genomic sequences and know DNA composition of organisms. The DNA sequencing is the discovery that uses the DNA composition to understand and decrypt the code to all biological life on earth as well as to understand and treat genetic diseases [2].

The appearance of sequencing technologies has played an important role in the analysis of genomic sequences of organisms. A DNA sequencer produces files containing DNA sequences [3]. These sequences are strings called reads on an alphabet formed by five letters {A, T, C, G, N}. The symbol N is used to represent an ambiguity. The first sequencing technologies were developed in 1977 by Sanger et al. [4] from Cambridge University awarded a Nobel Prize in chemistry in 1980 and Maxam et al. [5] from Harvard University. Their discovery opened the door to study the genetic code of living beings and brought their inspiration to researchers to the development of faster and efficient sequencing technology. Sanger sequencing has become the most applied technique of sequencing for its high efficiency and low radioactivity [6] and has been commercialized and automated as the "Sanger Sequencing Technology".

Sanger and Maxam-Gilbert sequencing technologies were the most common sequencing technologies used by biologists until the emergence of a new era of sequencing technologies opening new perspectives for genomes exploration and analysis. These sequencing technologies were firstly appeared by Roche's 454 technology in 2005 [7] and were commercialized as technologies capable of producing sequences with very high throughput and at much lower cost than the first sequencing technologies. These new sequencing technologies are generally known under the name of "Next Generation Sequencing (NGS) Technologies" or "High Throughput Sequencing Technologies".

NGS technologies produce a massively parallel analysis with a high-throughput from multiple samples at much reduced cost [8]. NGS technologies can be sequenced in parallel millions to billions of reads in a single run and the time required to generate the GigaBase sized reads is only a few days or hours making it best than the first generation sequencing such as Sanger sequencing. The human genome, for example, consists of 3 billion bps and is made up of DNA macromolecules of lengths varying from 33 to ~247 million bps, distributed in the 23 chromosomes located in each human cell nucleus, the sequencing of the human genome using the Sanger sequencing took almost 15 years, required the cooperation of many laboratories around the world and costed approximately 100 million US dollars, whereas the sequencing by NGS sequencers using the 454 Genome Sequencer FLX took two months and for approximately one hundredth of the cost [9]. Unfortunately, NGS are incapable to read the complete DNA sequence of the genome, they are limited to sequence small DNA fragments and generate millions of reads. This limit remains a negative point especially for genome assembly projects because it requires high computing resources.

NGS technologies continue to improve and the number of sequencers increases these last years. However, the literature divided NGS technologies into two types [3,10]. We distinguish the second generation sequencing technologies which refer to the newest sequencing technologies developed in the NGS environment after the first generation [11,12], they are characterized by the need to prepare amplified sequencing banks before starting the sequencing of amplified DNA clones [13] and there are the third generation sequencing technologies that are sequencing technologies recently appeared [6], in contrast to the second generation, these technologies are classified as Single Molecule Sequencing Technology [14] because they can make

sequencing a single molecule without the necessity to create the amplification libraries and that are capable of generating longer reads at much lower costs and in a shorter time.

Several previous reports and studies presented the sequencing technologies and detailed chemical mechanisms of each sequencing platform [10,11,13,15-23]. In the following, we present a brief review of the three existing generations of sequencing technologies (the first, second and third). We focus on sequencing methods and platforms characterizing each generation of sequencing (Table 1).

Platform	Instrument	Reads per run	Avg Read length (pb)	Read Type	Error Type	Error Rate (%)	Data Generated per run (Gb)	Year
First Generation								
ABI Sanger	3730xl	96	400 – 900*	SE	NA	0.3	0.00069 to 0.0021	2002
Second Generation								
454	GS20	200	100	SE, PE	indel	1	0.02	2005
454	GS FLX	400	250	SE, PE	indel	1	0.1	2007
454	GS FLX Titanium	1 M	450	SE, PE	indel	1	0.45	2009
454	GS FLX Titanium+	1 M	700	SE, PE	indel	1	0.7	2011
454	GS Junior	100	400	SE, PE	indel	1	0.04	2010
454	GS Junior+	100	700	SE, PE	indel	1	0.07	2014
Illumina	MiniSeq	25M (maximum)	150	SE, PE	mismatch	1	7.5 (maximum)	2013
Illumina	MiSeq	25M (maximum)	300	SE, PE	mismatch	0.1	15 (maximum)	2011
Illumina	NextSeq	400M (maximum)	150	SE, PE	mismatch	1	120 (maximum)	2014
Illumina	HiSeq	5B (maximum)	150	SE, PE	mismatch	0.1	1.5Tb (maximum)	2012
Illumina	HiSeq X	6B (maximum)	150	SE, PE	mismatch	0.1	1.8Tb (maximum)	2014
SOLiD	5500 W	3B	75	SE	mismatch	~0.1	160	2011
SOLiD	5500xl W	6B	75	SE	mismatch	~0.1	320	2013
Ion Torrent	PGM 314 chip v2	400.000-550.000	400	SE	indel	1	0.06 to 0.1	2011
Ion Torrent	PGM 316 chip v2	2M - 3M	200	SE	indel	1	0.6 to 1	2011
Ion Torrent	PGM 318 chip v2	4M - 5.5M	400	SE	indel	1	1.2 to 2	2013
Ion Torrent	Ion Proton	60M - 80M	200	SE	indel	1	10	2012
Ion Torrent	Ion S5/S5XL 520	3M - 5M	400	SE	indel	1	1.2 to 2	2015
Ion Torrent	Ion S5/S5XL 530	15M-20M	400	SE	indel	1	03 to 05	2015
Ion Torrent	Ion S5/S5XL 540	60M - 80M	400	SE	indel	1	NA	2015
Third Generation								
PacBio	RS C1	432	1300	SE	indel	15	0.54	2011
PacBio	RS C2	432	2500	SE	indel	15	0.5 to 1	2012
PacBio	RS C2 XL	432	4300	SE	indel	15	0.5 to 1	2012
PacBio	RS II C2 XL	564	4600	SE	indel	15	0.5 to 1	2013
PacBio	RS II P5 C3	528	8500	SE	indel	13	0.5 to 1	2014
PacBio	RS II P6 C4	660	13500	SE	indel	12	0.5 to 1	2014

PacBio	Sequel	350	10000	SE	NA	NA	7	2016
Oxford Nanopore	MinION Mk	100	9545	1D,2D	indel/mismatch	12	1.5	2015
Oxford Nanopore	PromethION	NA	9846	1D,2D	NA	NA	2Tb to 4Tb	2016
*depending on run module; NA: Not available; SE: Single End; PE: Paired End; M: Million; B: Billion; Gb: Gigabytes; Tb: Terabytes								

Table 1: Summary of NGS platforms and characteristics.

The First Generation of Sequencing

Sanger and Maxam-Gilbert sequencing technologies were classified as the First Generation Sequencing Technology [10,16] who initiated the field of DNA sequencing with their publication in 1977.

Sanger sequencing

Sanger Sequencing is known as the chain termination method or the dideoxynucleotide method or the sequencing by synthesis method. It consists in using one strand of the double stranded DNA as template to be sequenced. This sequencing is made using chemically modified nucleotides called dideoxy-nucleotides (dNTPs). These dNTPs are marked for each DNA bases by ddG, ddA, ddT, and ddC. The dideoxy-nucleotides are used dNTPs are used for elongation of nucleotide, once incorporated into the DNA strand they prevent the further elongation and the elongation is complete. Then, we obtain DNA fragments ended by a dNTP with different sizes. The fragments are separated according to their size using gel slab where the resultant bands corresponding to DNA fragments can be visualized by an imaging system (X-ray or UV light) [24,25]. Figure 1 details the Sanger sequencing technology.

The first genomes sequenced by the Sanger sequencing are phiX174 genome with size of 5374 bp [26] and in 1980 the bacteriophage λ genome with length of 48501 bp [27]. After years of improvement, Applied Biosystems is the first company that has automated Sanger sequencing. Applied Biosystems has built in 1995 an automatic sequencing machine called ABI Prism 370 based on capillary electrophoresis allowing fast an accurate sequencing. The Sanger sequencing was used in several sequencing projects of different plant species such as Arabidopsis [28], rice [29] and soybean [30] and the most emblematic achievement of this sequencing technology is the decoding of the first human genome [31].

The sanger sequencing was widely used for three decades and even today for single or low-throughput DNA sequencing, however, it is difficult to further improve the speed of analysis that does not allow the sequencing of complex genomes such as the plant species genomes and the sequencing was still extremely expensive and time consuming.

Maxam-Gilbert sequencing

Maxam-Gilbert is another sequencing belonging to the first generation of sequencing known as the chemical degradation method. Relies on the cleaving of nucleotides by chemicals and is most effective with small nucleotides polymers. Chemical treatment generates breaks at a small proportion of one or two of the four nucleotide bases in each of the four reactions (C, T+C, G, A+G). This reaction leads to a series of marked fragments that can be separated according to their size by electrophoresis [5,24].

The sequencing here is performed without DNA cloning. However, the development and improvement of the Sanger sequencing method favored the latter to the Maxam-Gilbert sequencing method, and it is also considered dangerous because it uses toxic and radioactive chemicals.

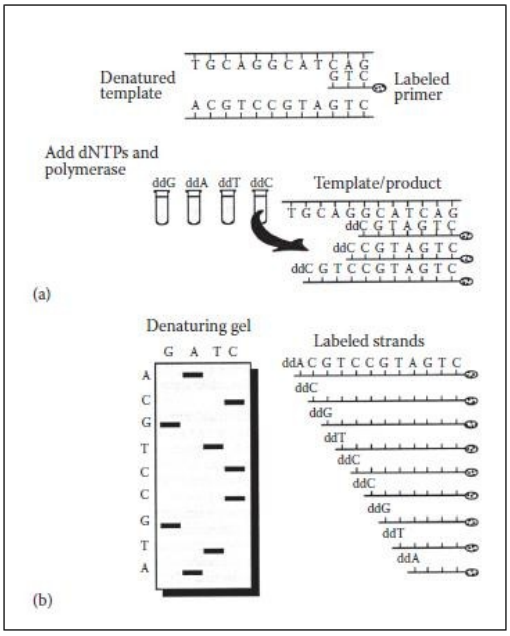


Figure 1: Sanger sequencing technology. (a) The sequencing reaction is performed by the presence of denatured DNA template, radioactively labeled primer, DNA polymerase, and dNTPs. The DNA polymerase is used to incorporate the dNTPs into the elongating DNA strand. Each of the four dNTPs is run in a separate reaction so the polymerization can randomly terminate at each base position. The end result of each reaction is a population of DNA fragments with different lengths, with the length of each fragment dependent on where the dNTPs is incorporated. (b) Illustrates the separation of these DNA fragments in a denaturing gel by electrophoresis. The radioactive labeling on the primer enables visualization of the fragments as bands on the gel. The bands on the gel represent the respective fragments shown to the right. The complement of the original template (read from bottom to top) is given on the left margin of the sequencing gel. (From P Moran, Overview of commonly used DNA techniques, in LK Park, P Moran, and RS Waples, eds., Application of DNA Technology to the Management of Pacific Salmon, 1994, 15–26, Department of Commerce, NOAA Technical Memorandum NMFS-NWFSC-17. © Paul Moran, NOAA's Northwest Fisheries Science Center. With permission).

The Second Generation of Sequencing

The first generation of sequencing was dominant for three decades especially Sanger sequencing, however, the cost and time was a major stumbling block. In 2005 and in subsequent years, have marked the emergence of a new generation of sequencers to break the limitations of the first generation. The basic characteristics of second generation sequencing technology are: (1) The generation of many millions of short reads in parallel, (2) The speed up of sequencing the process compared to the first generation, (3) The low cost of sequencing and (4) The sequencing output is directly detected without the need for electrophoresis.

Short read sequencing approaches divided under two wide approaches: sequencing by ligation (SBL) and sequencing by synthesis (SBS), (more details for these sequencing categories are presented in [22,32]) and are mainly classified into three major sequencing platforms: Roche/454 launched in 2005, Illumina/Solexa in 2006 and in 2007 the ABI/SOLiD. We will briefly describe these commonly utilized sequencing platforms.

Roche/454 sequencing

Roche/454 sequencing appeared on the market in 2005, using pyrosequencing technique which is based on the detection of pyrophosphate released after each nucleotide incorporation in the new synthetic DNA strand (http://www.454.com). The pyrosequencing technique is a sequencing-by-synthesis approach.

DNA samples are randomly fragmented and each fragment is attached to a bead whose surface carries primers that have oligonucleotides complementary to the DNA fragments so each bead is associated with a single fragment (Figure 2A). Then, each bead is isolated and amplified using PCR emulsion which produces about one million copies of each DNA fragment on the surface of the bead (Figure 2B). The beads are then transferred to a plate containing many wells called picotiter plate (PTP) and the pyrosequencing technique is applied which consists in activating of a series of downstream reactions producing light at each incorporation of nucleotide. By detecting the light emission after each incorporation of nucleotide, the sequence of the DNA fragment is deduced (Figure 2C) [15]. The use of the picotiter plate allows hundreds of thousands of reactions occur in parallel, considerably increasing sequencing throughput [14]. The latest instrument launched by Roche/454 called GS FLX+ that generates reads with lengths of up to 1000 bp and can produce ~1Million reads per run (454.com GS FLX+Systems http://454.com/products/gs-flx-system/index.asp). Other characteristics of Roche/454 instruments are listed in [16,25].

The Rche/454 is able to generate relatively long reads which are easier to map to a reference genome. The main errors detected of sequencing are insertions and deletions due to the presence of homopolymer regions [33,34]. Indeed, the identification of the size of homopolymers should be determined by the intensity of the light emitted by pyrosequencing. Signals with too high or too low intensity lead to under or overestimation of the number of nucleotides which causes errors of nucleotides identification.

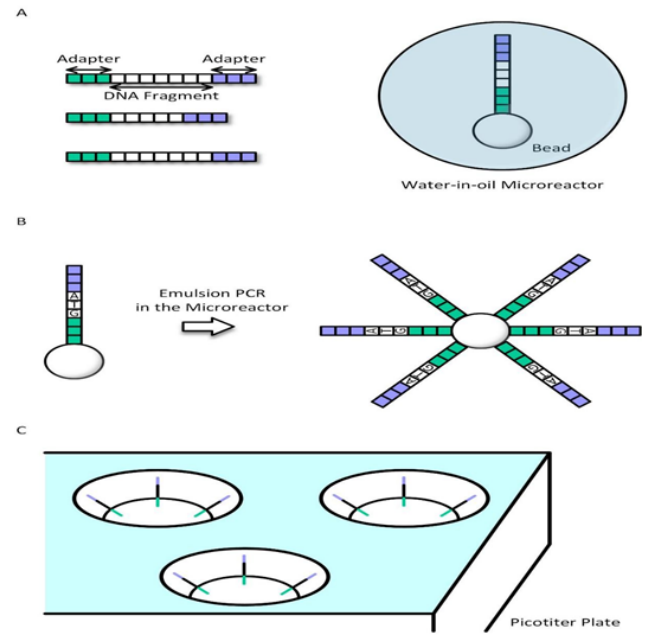


Figure 2: Roche/454 sequencing technology [39].

Ion torrent sequencing

Life Technologies commercialized the Ion Torrent semiconductor sequencing technology in 2010 (https://www.thermofisher.com/us/en/home/brands/ion-torrent.html). It is similar to 454 pyrosequencing technology but it does not use fluorescent labeled nucleotides like other second-generation technologies. It is based on the detection of the hydrogen ion released during the sequencing process [35].

Specifically, Ion Torrent uses a chip that contains a set of micro wells and each has a bead with several identical fragments. The incorporation of each nucleotide with a fragment in the pearl, a hydrogen ion is released which change the pH of the solution. This change is detected by a sensor attached to the bottom of the micro well and converted into a voltage signal which is proportional to the number of nucleotides incorporated (Figure 3).

The Ion Torrent sequencers are capable of producing reads lengths of 200 bp, 400 bp and 600 bp with throughput that can reach 10 Gb for ion proton sequencer. The major advantages of this sequencing technology are focused on read lengths which are longer to other SGS sequencers and fast sequencing time between 2 and 8 hours. The major disadvantage is the difficulty of interpreting the homopolymer sequences (more than 6 bp) [21,36] which causes insertion and deletion (indel) error with a rate about ~1%.

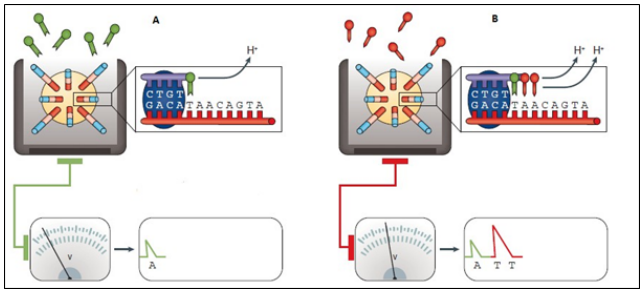


Figure 3: Ion torrent sequencing technology [22].

Illumina/Solexa sequencing

The Solexa company has developed a new method of sequencing. Illumina company (http://www.illumina.com) purchased Solexa that started to commercialize the sequencer Ilumina/Solexa Genome Analyzer (GA) [3,37]. Illumina technology is sequencing by synthesis approach and is currently the most used technology in the NGS market.

The sequencing process is shown in Figure 4. During the first step, the DNA samples are randomly fragmented into sequences and adapters are ligated to both ends of each sequence. Then, these adapters are fixed themselves to the respective complementary adapters, the latter are hooked on a slide with many variants of adapters (complementary) placed on a solid plate (Figure 4A). During the second step, each attached sequence to the solid plate is amplified by "PCR bridge amplification" that creates several identical copies of each sequence; a set of sequences made from the same original sequence is called a cluster. Each cluster contains approximately one million copies of the same original sequence (Figure 4B). The last step is to determine each nucleotide in the sequences, Illumina uses the sequencing by synthesis approach that employs reversible terminators [38] in which the four modified nucleotides, sequencing primers and DNA polymerases are added as a mix, and the primers are hybridized to the sequences. Then, polymerases are used to extend the primers using the modified nucleotides. Each type of nucleotide is labeled with a fluorescent specific in order for each type to be unique. The nucleotides have an inactive 3'-hydroxyl group which ensures that only one nucleotide is incorporated. Clusters are excited by laser for emitting a light signal specific to each nucleotide, which will be detected by a coupled-charge device (CCD) camera and Computer programs will translate these signals into a nucleotide sequence (Figure 4C). The process continues with the elimination of the terminator with the fluorescent label and the starting of a new cycle with a new incorporation [21,39].

The first sequencers Illumina/Solexa GA has been able to produce very short reads ~35 bp and they had an advantage in that they could produce paired-end (PE) short reads, in which the sequence at both ends of each DNA cluster is recorded. The output data of the last Illumina sequencers is currently higher than 600 Gpb and lengths of short reads are about 125 bp. Details on Illumina sequencers [13].

One of the main drawbacks of the Illumina/Solexa platform is the high requirement for sample loading control because overloading can result in overlapping clusters and poor sequencing quality. The overall error rate of this sequencing technology is about 1%. Substitutions of nucleotides are the most common type of errors in this technology [40], the main source of error is due to the bad identification of the incorporated nucleotide.

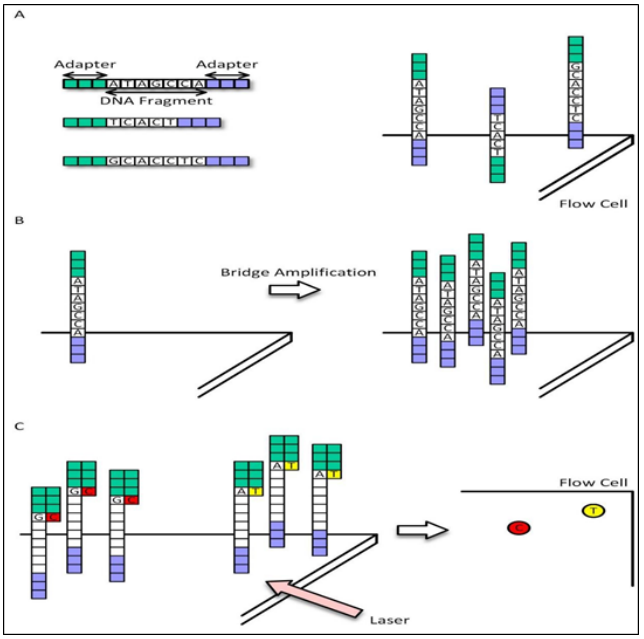


Figure 4: Illumina sequencing technology [39].

ABI/SOLiD sequencing

Supported Oligonucleotide Ligation and Detection (SOLiD) is a NGS sequencer Marketed by Life Technologies (http://www.lifetechnologies.com). In 2007, Applied Biosystems (ABI) has acquired SOLiD and developed ABI/SOLID sequencing technology that adopts by ligation (SBL) approach [3].

The ABI/SOLiD process consists of multiple sequencing rounds. It starts by attaching adapters to the DNA fragments, fixed on beads and cloned by PCR emulsion. These beads are then placed on a glass slide and the 8-mer with a fluorescent label at the end are sequentially ligated to DNA fragments, and the color emitted by the label is recorded (Figure 5A). Then, the output format is color space which is the encoded form of the nucleotide where four fluorescent colors are used to represent 16 possible combinations of two bases. The sequencer repeats this ligation cycle and each cycle the complementary strand is removed and a new sequencing cycle starts at the position n-1 of the template. The cycle is repeated until each base is sequenced twice (Figure 5B). The recovered data from the color space can be translated to letters of DNA bases and the sequence of the DNA fragment can be deduced [15].

ABI/SOLiD launched the first sequencer that produce short reads with length 35 bp and output of 3 Gb/run and continued to improve their sequencing which increased the length of reads to 75 bp with an output up to 30 Gb/run [22,23]. The strength of ABI/SOLiD platform is high accuracy because each base is read twice while the drawback is the relatively short reads and long run times. The errors of sequencing in this technology is due to noise during the ligation cycle which causes error identification of bases. The main type of error is substitution.

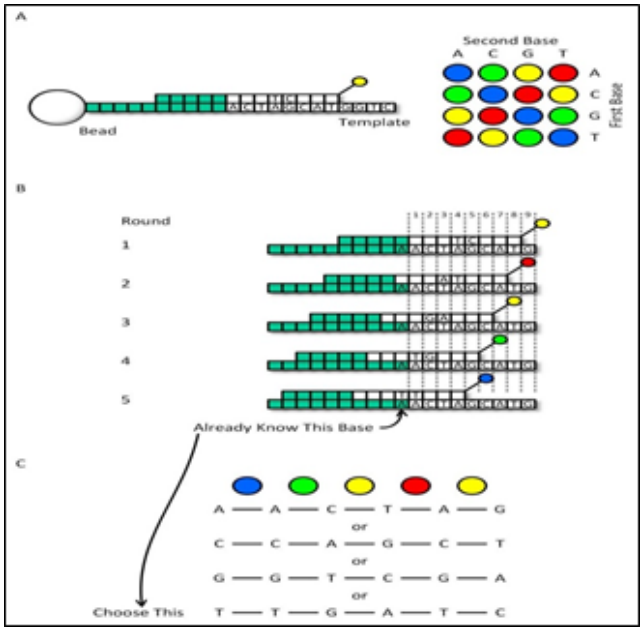


Figure 5: ABI/SOLID sequencing technology [39].

The Third Generation of Sequencing

The second-generation of sequencing technologies previously discussed have revolutionized the analysis of DNA and have been the most widely used compared to the first generation of sequencing technologies. However, the SGS technologies generally require PCR amplification step which is a long procedure in execution time and expansive in sequencing price. Also, it became clear that the genomes are very complex with many repetitive areas that SGS technologies are incapable to solve them and the relatively short reads made genome assembly more difficult. To remedy the problems caused by SGS technologies, scientists have developed a new generation of sequencing called "third generation sequencing". These third generations of sequencing have the ability to offer a low sequencing cost and easy sample preparation without the need PCR amplification in an execution time significantly faster than SGS technologies. In addition, TGS are able to produce long reads exceeding several kilobases for the resolution of the assembly problem and repetitive regions of complex genomes.

There are two main approaches that characterize TGS [22]: The single molecule real time sequencing approach (SMRT) [38] that was developed by Quake laboratory [41-43] and the synthetic approach that rely on existing short reads technologies used by Illumina (Moleculo) [43] and 10xGenomics (https://www.10xgenomics.com) to construct long reads. The most widely used TGS technology approach is SMRT and the sequencers that have used this approach are Pacific Biosciences and Oxford Nanopore sequencing (specifically the MinION sequencer).

In the following, we present the two most widely used sequencing platforms in TGS to know Pacific Biosciences and the MinION sequencing from Oxford Nanopore technology.

Pacific biosciences SMRT sequencing

Pacific Biosciences (http://www.pacificbiosciences.com/) developed the first genomic sequencer using SMRT approach and it's the most widely used third-generation sequencing technology.

Pacific Biosciences uses the same fluorescent labelling as the other technologies, but instead of executing cycles of amplification nucleotide, it detects the signals in real time, as they are emitted when the incorporations occur. It uses a structure composed of many SMRT cells, each cell contains microfabricated nanostructures called zero-mode waveguides (ZMWs) which are wells of tens of nanometers in diameter microfabricated in a metal film which is in turn deposited onto a glass substrate [44,45]. These ZMWs exploit the properties of light passing through openings with a diameter less than its wavelength, so light cannot be propagated. Due to their small diameter, the light intensity decreases along the wells and the bottom of the wells illuminated (Figure 6A). Each ZMW contains a DNA polymerase attached to their bottom and the target DNA fragment for sequencing. During the sequencing reaction, the DNA fragment is incorporated by the DNA polymerase with fluorescent labeled nucleotides (with different colors). Whenever a nucleotide is incorporated, it releases a luminous signal that is recorded by sensors (Figure 6B). The detection of the labeled nucleotides makes it possible to determine the DNA sequence.

Compared to SGS, Pacific Bioscience technology has several advantages. The preparation of the sample is very fast, it takes 4 to 6 hours instead of days [16]. In addition, the long-read lengths, currently averaging ~10 kbp [46] but individual very long reads can be as long as 60 kbp, which is longer than that of any SGS technology. Pacific Biosciences sequencing platforms have a high error rate of about 13% [13] dominated by insertions and deletions errors. These errors are randomly distributed along the long read [47].

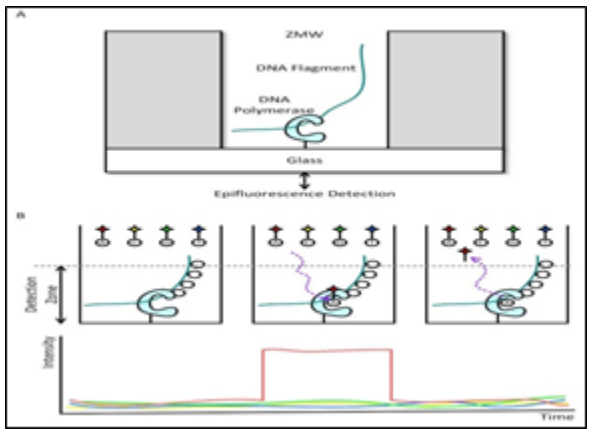


Figure 6: Pacific biosciences sequencing technology [39].

Oxford nanopore sequencing

The Oxford Nanopore sequencing (ONT) was developed as a technique to determine the order of nucleotides in a DNA sequence. In 2014, Oxford Nanopore Technologies released the MinION [48] device that promises to generate longer reads that will ensure a better resolution structural genomic variants and repeat content [49]. It's a mobile single-molecule Nanopore sequencing measures four inches in

length and is connected by a USB 3.0 port of a laptop computer. This device has been released for testing by a community of users as part of the MinION Access Program (MAP) to examine the performance of the MinION sequencer [50].

In this sequencing technology, the first strand of a DNA molecule is linked by a hairpin to its complementary strand. The DNA fragment is passed through a protein nanopore (a nanopore is a nanoscale hole made of proteins or synthetic materials [39]). When the DNA fragment is translated through the pore by the action of a motor protein attached to the pore, it generates a variation of an ionic current caused by differences in the moving nucleotides occupying the pore (Figure 7A). This variation of ionic current is recorded progressively on a graphic model and then interpreted to identify the sequence (Figure 7B). The sequencing is made on the direct strand generating the "template read" and then the hairpin structure is read followed by the inverse strand generating the "complement read", these reads is called "1D". If the "temple" and "complement" reads are combined, then we have a resulting consensus sequence called "two direction read" or "2D" [51,52].

Among the advantages offered by this sequencer: first, it's low cost and small size. Then, the sample is loaded into a port on the device and data is displayed on the screen and generated without having to wait till the run is complete. And, MinION can provide very long reads exceeding 150 kbp which can improve the contiguity of the denovo assembly. However, MinION produces a high error rate of ~12% distributed about ~3% mismatchs, ~4% insertions and ~5% deletions [53].

The ONT technology has continued to evolve. Recently, a new instrument has emerged called "PromethION"[54]; it is the bigger brother of the MinION [55]. It is an autonomous worktable sequencer with 48 individual flow cells each with 3000 pores (equivalent to 48 MinIONs) operating at 500 bp [51] per second which is sufficiently powerful to achieve an ultra-high throughput needed for sequencing large genomes such as the human genome. Although the PromethION is not commercially available, the ONT announces that it is capable of producing ~2 to 4 Tb for a duration of 2 days and a length of reads [22] which can attain 200 Kpb which puts this sequencer in competition with the PacBioRSII sequencer from pacific biosciences in terms of read length and HiSeq sequencer from Illumina in cost.

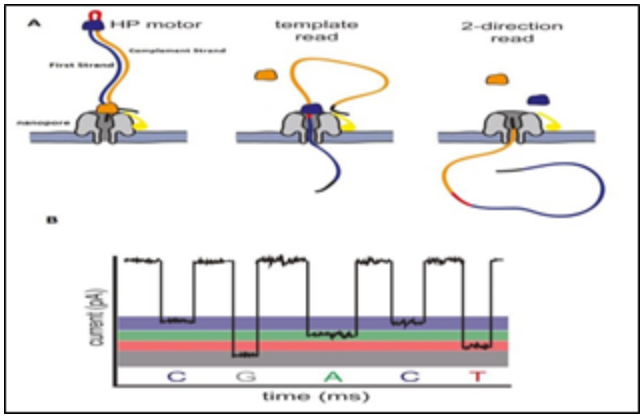


Figure 7: Oxford nanopore MinION sequencing [21].

Conclusion

The first method of sequencing came about half a century ago, and since then, sequencing technologies have continued to evolve especially after the appearance of the first sequencers from NGS technology which appeared in 2005. These technologies are characterized by their high throughput which gives the opportunity to produce millions of reads with inexpensive sequencing. NGS technologies are now the starting point for several areas of research based on the study and analysis of biological sequences.

In this review, we presented a concise overview of the generations of sequencing technologies by beginning with the first-generation sequencing history followed by the main commonly used NGS platforms. Nevertheless, there are significant challenges in NGS technologies, including the difficulty of storing and analyzing the data generated by these technologies. This is mainly due to the production of a high number of reads. In the coming years, new sequencing platforms will appear producing a larger amount of data (in Terabyte) which requires the development of new approaches and applications capable of analyzing this large amount of data.

Conflict of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

References

1. Watson JD, Crick FH (1953) Molecular structure of nucleic acids: A structure for deoxyribose nucleic acid. Nature 171: 737-738.
2. Le Tourneau, Christophe, Kamal, Maud (2015) Pan-cancer integrative molecular portrait towards a new paradigm in precision medicine. Springer.
3. Shendure J, Ji H (2008) Next-generation DNA sequencing. Nature Biotechnology 26: 135–1145.
4. Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci 74: 5463–5467.
5. Maxam AM, Gilbert WA (1977) A new method for sequencing DNA. Proc Natl Acad Sci 74: 560-564.
6. Pareek CS, Smoczynski R, Tretyn A (2011) Sequencing technologies and genome sequencing. J Appl Genet 52: 413–35.
7. Qiang-long Z, Shi L, Peng G, Fei-shi L (2014) High-throughput sequencing technology and its application. Journal of Northeast Agricultural University 21: 84-96.
8. Mardis ER (2011) A decade's perspective on DNA sequencing technology. Nature 470: 198-203.
9. Wheeler DA, Srinivasan M, Egholm M, Shen Y, Chen L, et al. (2008) The complete genome of an individual by massively parallel DNA sequencing. Nature 452: 872-876.
10. Thudi M, Li Y, Jackson SA, May GD, Varshney RK (2012) Current state-of-art of sequencing technologies for plant genomics research. Brief Funct Genomics 11: 3-11.
11. Michael LM (2010) Sequencing technologies – the next generation. Nature Reviews Genetics 11: 31-46.
12. Schatz MC, Delcher AL, Salzberg SL (2010) Assembly of large genomes using second generation sequencing. Genome Research 20: 1165- 1173.
13. Kulski JK (2016) Next-generation sequencing-An overview of the history, tools, and "Omic" applications, next generation sequencing-advances, applications and challenges. InTech.
14. Vezzi F (2012) Next generation sequencing revolution challenges: Search, assemble, and validate genomes. Ph.D, Universita degli Studi di Udine, Italy.

15. Mardis ER (2008) Next-generation DNA sequencing methods. Annu Rev Genomics Hum Genet 9: 387-402.
16. Liu L, Li Y, Li S, Hu N, He Y, et al. (2012) Comparison of next-generation sequencing Systems. Journal of Biomedicine and Biotechnology.
17. Guzvic M (2013) The history of DNA sequencing. J Med Biochem 32: 301-12.
18. Hui P (2012) Next generation sequencing: Chemistry, technology and applications. Top Curr Chem 336: 1-18.
19. van Dijk EL, Auger H, Jaszczyszyn Y, Thermes C (2014) Ten years of next generation sequencing technology. Trends in Genetics 30: 418-26.
20. Heather JM, Chain B (2015) The sequence of sequencers: The history of sequencing DNA. Genomics 107: 1-8.
21. Reuter JA, Spacek DV, Snyder MP (2015) High-throughput sequencing technologies. Molecular Cell 58: 586-597.
22. Goodwin S, McPherson JD, Richard McCombie W (2016) Coming of age: Ten years of next-generation sequencing technologies. Nature Reviews Genetics 17: 333–351.
23. Alic AS, Ruzafa D, Dopazo J, Blanquer I (2016) Objective review of de novo stand-alone error correction methods for NGS data. In: WIREs Computational Molecular Science.
24. Masoudi-Nejad A, Narimani Z, Hosseinkhan N (2013) Next generation sequencing and sequence assembly. Methodologies and algorithms. Springer.
25. El-Metwally S, Ouda OM, Helmy M (2014) Next generation sequencing technologies and challenges in sequence assembly. Springer.
26. Sanger F, Coulson AR (1975) A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. J Mol Biol 94: 441-448.
27. Sanger F, Coulson AR, Barrell BG, Smith AJ, Roe BA (1980) Cloning in single stranded bacteriophage as an aid to rapid dna sequencing. J Mol Biol 143: 161-178.
28. The Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. Nature 408: 796-815.
29. Goff SA, Ricke D, Lan TH, Presting G, Wang R, et al. (2002) A draft sequence of the rice genome (Oryza sativa L. ssp. japonica). Science 296: 92-100.
30. Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, et al. (2010) Genome sequence of the paleopolyploid soybean. Nature 463: 178-83.
31. Durbin RM (2010) A map of human genome variation from population-scale sequencing. Nature 467: 1061-73.
32. Myllykangas S, Buenrostro J, Ji HP (2012) Overview of sequencing technology platforms, bioinformatics for high throughput sequencing.
33. Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, et al. (2005) Genome sequencing in microfabricated high-density picolitre reactors. Nature 437: 376-80.
34. Huse SM, Huber JA, Morrison HG, Sogin ML, Welch DM (2007) Accuracy and quality of massively parallel DNA pyrosequencing. Genome Biol 8: R143.
35. Rotheberg JM, Hinz W, Rearrick TM, Schultz J, Mileski W, et al. (2011) An integrated semiconductor device enabling non-optical genome sequencing. Nature 475: 348-52.
36. Loman NJ, Misra RV, Dallman TJ, Constantinidou C, Gharbia SE, et al. (2012) Performance comparison of benchtop high-throughput sequencing platforms. Nature Biotechnol 30: 434-439.
37. Balasubramanian S (2015) Solexa sequencing: Decoding genomes on a population scale. Clin Chem 61: 21-24.
38. Bentley DR, Balasubramanian S, Swerdlow HP, Smith GP, Milton J, et al. (2008) Accurate whole human genome sequencing using reversible terminator chemistry. Nature 456: 53–59.
39. Heo, Yun (2015) Improving quality of high-throughput sequencing reads.
40. Dohm JC, Lottaz C, Borodina T, Himmelbauer H (2008) Substantial biases in ultra-short read data sets from high-throughput DNA sequencing. Nucleic Acids Res 36: e105.
41. Eid J, Fehr A, Gray J (2009) Real-time DNA sequencing from single polymerase molecules. Science 323: 133-138.
42. Braslavsky I, Hebert B, Kartalov E, Quake SR (2003) Sequence information can be obtained from single DNA molecules. Proceedings of the National Academy of Sciences of the USA 100: 3960-3964.
43. Harris TD, Buzby PR, Babcock H, Beer E, Bowers J, et al. (2008) Single-molecule DNA sequencing of a viral genome. Science 320: 106-9.
44. McCoy RC, Taylor RW, Blauwkamp TA, Kelley JL, Kertesz M, et al. (2014) Illumina TruSeq synthetic long-reads empower de novo assembly and resolve complex, highly-repetitive transposable elements. PLoS ONE 9: e106689.
45. Rhoads A, Au KF (2015) PacBio sequencing and its applications. Genomics, Proteomics & Bioinformatics 13: 178-289.
46. Chin CS, Peluso P, Sedlazeck FJ, Nattestad M, Concepcion GT, et al. (2016) Phased diploid genome assembly with single molecule real- time sequencing. BioRxiv 13: 1050-1054.
47. Koren S, Schatz M, Walenz B, Martin J, Howard J, et al. (2012) Hybrid error correction and de novo assembly of single-molecule sequencing reads. Nature Biotechnology 30: 693-700.
48. Mikheyev AS, Tin MMY (2014) A first look at the oxford nanopore MinION sequencer. Molecular Ecology Resources 14: 1097–1102.
49. Laehnemann D, Borkhardt A, McHardy AC (2015) Denoising DNA deep sequencing data—high-throughput sequencing errors and their correction. Brief Bioinformatics 17: 154-79.
50. Laver T, Harrisona J, O'Neill PA, Moorea K, Farbosa A, et al. (2015) Studholme. Assessing the performance of the oxford nanopore technologies MinION. Biomolecular Detection and Quantification 3: 1-8.
51. Lu H, Giordano F, Ning Z (2016) Oxford nanopore MinION sequencing and genome assembly. Genomics Proteomics Bioinformatics 14: 265-279.
52. Jain M, Hugh EO, Paten B, Akeson M (2016) The oxford nanopore MinION: Delivery of nanopore sequencing to the genomics community. Genome Biology 17: 239.
53. Ip CL, Loose M, Tyson JR, de Cesare M, Brown BL, et al. (2015) MinION analysis and reference consortium: Phase 1 data release and analysis. F1000Research 4: 1075.
54. Karow J (2014) Oxford Nanopore presents details on new high-throughput sequencer, improvements to MinIon.
55. http://www.opiniomics.org/doing-the-maths-on-promethion-throughput