

We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

4,800

Open access books available

122,000

International authors and editors

135M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.

For more information visit www.intechopen.com



High-Throughput Sequencing and Metagenomic Data Analysis

*Ahmed Shuikan, Sulaiman Ali Alharbi,
Dalal Hussien M. Alkhalifah and Wael N. Hozzein*

Abstract

Metagenomic approaches are a growing branch of science and have many applications in different fields. Metagenomics seems to be the ideal culture-independent technique for unraveling the biodiversity of soils and to study how this biodiversity is affected with continuously changing conditions. In addition, its application in clinical and diagnostic approaches was reported. The emergence of several next-generation sequencing (NGS) strategies enriched the metagenomics. The combination between NGS and metagenomic approaches helped the investigators resolve several issues regarding the microbial diversity and the functions and relationships among different microbial flora. A number of NGS approaches were developed including Roche/454 pyrosequencing, Illumina/Solexa sequencing, and Applied Biosystems/SOLiD sequencing. In this chapter, different NGS platforms are discussed in terms of principle, advantages, and limitations. In addition, third-generation sequencing technologies are also addressed.

Keywords: high throughput, metagenomics workflow, sequencing approaches, metagenomic data analysis

1. Introduction

The development of next-generation sequencing (NGS) techniques provides high-throughput sequence analysis with the ability to simultaneously and independently sequence billions of DNA molecules. The combination between such technologies and metagenomic approaches helped the investigators study the microbial diversity and understand the functions and relationships among different microbial flora [1]. The use of metagenomic NGS by microbiologists overcomes several limitations and secured the unbiased methods to study the microbial flora in any given environment [2]. Thus, the dynamic of complex communities particularly those with non-cultivable microorganisms can be resolved [3, 4]. In addition, metagenomic NGS found its way in the field of clinical and diagnostic approaches [5, 6]. In the clinical field, NGS was used to inform the real-time incidence and prevention response to human parainfluenza 3 virus infections [7] and for cerebrospinal fluid diagnostics [8].

Several NGS platforms were developed since 2006 with numerous applications in genetic and biological research fields. Of these platforms, the most commonly used include Roche/454 pyrosequencing, Illumina/Solexa sequencing, and Applied Biosystems/SOLiD sequencing. The principle of all these NGS depends on the detection of luminescent signals released by the base incorporation during the

sequencing process [4]. They also share the same workflow which in the order, DNA extraction, library construction, DNA template preparation, and automated sequence analysis [9]. In this chapter, different NGS platforms are discussed in terms of principle, advantages, and limitations. In addition, third-generation sequencing technologies are also addressed.

2. Workflow of metagenomics

2.1 The sampling process and library construction for metagenomic analysis

Metagenomic analysis is a sophisticated process and involves several steps. Of these steps, the sampling process is very crucial for the downstream applications. Sample collection, preparation, and storage should be handled carefully to prevent lysis and decomposition of the sample compositions. Multiple freezing–thawing cycles may cause changes in the microbial community profile under investigation [10]. As well, a suitable DNA extraction protocol should be adopted to cope with the different chemical and physical characteristics of each sample. For instance, soils contain many substances that are co-extracted with the genomic DNA and may have inhibitory effects on the downstream experiments. Examples include humic and fulvic acids [11]. Therefore, optimization and comparison between different extraction methods are usually required for each type of samples [12, 13, 14, 15].

The extracted DNA is used to construct the DNA library. This is usually achieved by connecting specific adaptors to one or both ends of the DNA fragments [16]. The reason for utilizing DNA adaptor is to deal with the pool of samples and then connect them to its original sample. Handling DNA at this stage should be careful to avoid chemical, physical, or enzymatic damage of DNA molecules [17]. The construction of a DNA library is usually achieved through two approaches. The first one is called meta-pair where the library is characterized by long fragment insert. The second approach is called paired-end libraries with short fragment insert. In both approaches, the DNA is fragmented into different fragment sizes that would allow for their cloning. The DNA fragments obtained from such processes are cloned into the proper cloning vector. The size of the resulting fragments determines the suitable vector for the cloning process. The small DNA fragments are usually cloned into plasmid vectors, whereas fragments up to 40 kbp are cloned into cosmid or fosmid vectors. Bacterial artificial chromosome (BAC) vectors are usually used to clone inserts with sizes that exceed 40 Kbp [18]. Finally, the free adaptor, dimers of the adaptor, and any other artifacts must be removed to avoid noisy sequencing data [17].

2.2 Sequencing approaches

During the 1970s, the first-generation sequencing techniques, chain termination [19], and chemical sequencing approaches [20] were developed. In contrast to the chemical sequencing approach, the Sanger sequencing method ultimately prevailed and found immense applications due to its simplicity and is more amenable to being scaled up [21]. Simply, the basis of Sanger sequencing depends on the incubation of a specific primer and the template DNA in the presence of DNA polymerase. The reaction is accomplished by the addition of a mixture of deoxyribonucleotide triphosphates and dNTPs' dideoxyribonucleotide triphosphates for chain termination, one of which was labeled with phosphorus-32. The resulting pool of DNA amplicons will be with the same 5' residue and different dNTP residues at the 3' end (**Figure 1**). This pool of DNA fragments is then fractionated by denaturing polyacrylamide gel electrophoresis giving a band pattern. In this way, DNA decoding can be achieved

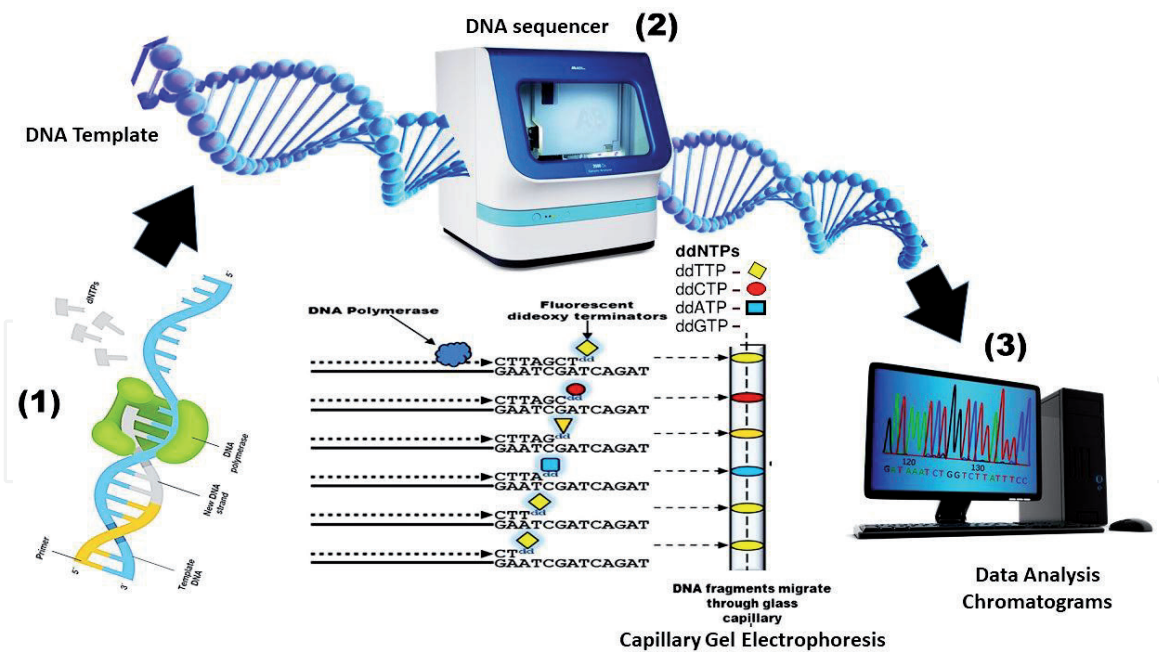


Figure 1. Sanger DNA sequencing. (1) The gene to be decoded is amplified by PCR. (2) The sequencing process is performed by the addition of modified 2',3'-dideoxynucleotide (ddNTPs) to the nascent chain. The modified nucleotides act by terminating the chain extension, and the resulting DNA fragments of different sizes are eluted by capillary gel electrophoresis. (3) Chromatograms are then analyzed to obtain the DNA sequences.

by the use of nucleotide analogs and other nucleotides in separate incubations and concomitant electrophoretic analysis [22]. Currently, the use of fluorescent dNTPs associated with the capillary electrophoresis provides full automation of the Sanger approach. This modification allows retrieving up to 96 sequences per run with an average 800–1000 bp size of DNA fragments [21, 23, 24]. Although the Sanger sequencing was the mainstay of the original human genome project, this approach still has some limitations. These limitations include high cost and low throughput, and it is inadequate for studying unculturable organisms in complex environments [25].

2.2.1 Next-generation sequencing (NGS)

Due to the limitations of Sanger sequencing technique, next-generation sequencing emerged in 2005 [26]. Indeed, next-generation sequencing has made it possible to study and identify organisms directly from their habitats without prior preparations [27]. Compared to the first-generation sequencing, NGS can generate several hundred thousand to millions of sequencing reads in parallel. As well, sequencing can be generated without some conventional steps such as vector-based cloning procedure and hence reduces the chance of DNA contamination from other organisms [28]. Therefore, several next-generation sequencing platforms have been introduced including Roche 454, Illumina®, Applied Biosystems SOLiD sequencer, and Ion Torrent. All next-generation sequencing or real-time sequencing (Roche 454, Illumina®, and AB SOLiD) utilized optical sensors that detect luminescent signal, which are produced during incorporation of bases in the sequence. The principles and characteristics of NSG, SGS, and TGS are summarized in **Table 1** [21]. In the subsequent sections, the features and limitations of each of the NGS techniques are discussed.

2.2.1.1 Roche 454 genome sequence

Roche/454 pyrosequencing is the first NGS technology that launched and became commercially available in 2005. It uses real-time sequencing-by-synthesis

	First generation	Second generation	Third generation
Fundamental technology	Size separation of specifically end-labeled DNA fragments, produced by SBS or degradation	Wash-and-scan SBS	SBS, by degradation, or direct physical inspection of the DNA molecule
Resolution	Averaged across many copies of the DNA molecule being sequenced	Averaged across many copies of the DNA molecule being sequenced	Single-molecule resolution
Current raw read accuracy	High	High	Moderate
Current read length	Moderate (800–1000 bp)	Short, generally much shorter than Sanger sequencing	Long, 1000 bp, and longer in commercial systems
Current throughput	Low	High	Moderate
Current cost	High cost per base	Low cost per base	Low-to-moderate cost per base
RNA sequencing method	cDNA sequencing	cDNA sequencing	Direct RNA sequencing and cDNA sequencing
Time from start of sequencing reaction to result	Hours	Days	Hours
Sample preparation	Moderately complex, PCR amplification not required	Complex, PCR amplification required	Ranges from complex to very simple depending on technology
Data analysis	Routine	Complex because of large data volumes and because short reads complicate assembly and alignment algorithms	Complex because of large data volumes and because technologies yield new types of information and new signal processing challenges
Primary results	Base calls with quality values	Base calls with quality values	Base calls with quality values, potentially other base information such as kinetics

Table 1.
The features and principles of first-generation sequencing, SGS, and TGS.

(SBS) pyrosequencing technology, and it depends on the detection of pyrophosphate (PPi) molecule that is initiated from the incorporation of a nucleotide in the DNA polymerase (**Figure 2**) [29]. Briefly, the 454 pyrosequencing technology is proceeding as follows: (i) the library fragments are connected to beads that carry oligonucleotides complementary to adapter sequence ligated at the ends, (ii) amplifying the library fragments by emulsion PCR resulting in DNA beads that carry millions of copies of DNA fragments on their surface, and (iii) the amplified beads are inserted into picotiter plate (PTP) that consists of millions of wells. Each well can hold only one amplified bead and contains diluted pyrosequence enzyme beads, DNA amplified beads, PPiase beads, and pyrosequence beads. Finally, the light emission from PTP is recorded by a CCD camera and is translated to nucleotide sequences [29]. In comparison with other NGS platform, 454 pyrosequencing has

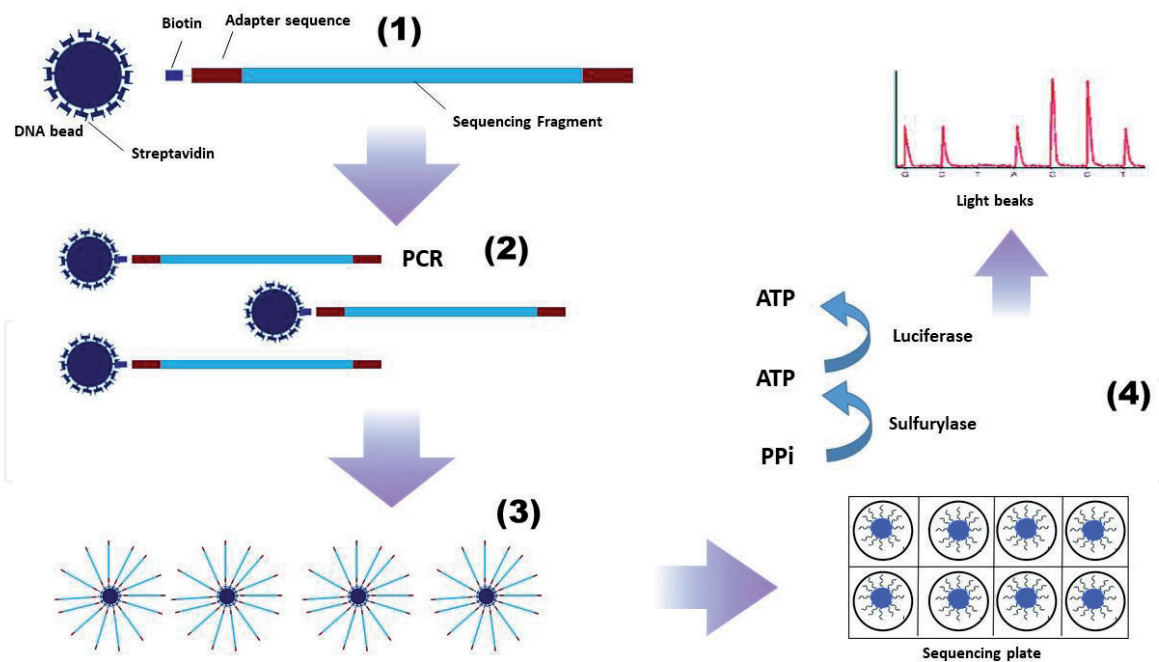


Figure 2. Pyrosequencing technique. (1) Beads coated with either streptavidin or complementary oligonucleotides complementary to adapter sequences attached to the ends of the fragment to be sequenced. This allows the binding of sequencing fragments to the beads. (2) The fragments to be sequenced are amplified through emulsion PCR. (3) Loaded beads are transferred into the sequencing plate with millions of wells. (4) By the addition of a nucleotide to the nascent chain that is connected to the beads by DNA polymerase, the ATP sulfurylase enzyme converts released pyrophosphate to ATP with the emission of light that is detected by a CCD camera and is translated to nucleotide sequences.

the longest reading (up to 1000–1200 bp). On the other hand, 454 pyrosequencing has the highest cost per base and the lowest output [30].

2.2.1.2 Illumina sequencing (Solexa genome analyzer)

Illumina, formerly known as Solexa, has been introduced commercially in 2007. Illumina technology utilizes bridge PCR amplification coupled with SBS in the flow cell (**Figure 3**). Simply, the principle of Illumina sequencing is that the DNA fragments with barcoding primer (adaptor) are attached to the flow cell. The sequencing reaction is performed in the flow cell by adding labeled nucleotides. When the nucleotide is incorporated, a luminescent signal is generated and then recorded by optical sensors. After that, the fluorescent molecules are removed and the next labeled nucleotide incorporated. However, the DNA fragment can be sequenced on one side that is called single-end (SE) or from both sides known as paired-end (PE). Nowadays, the most common sequencing used is PE due to the ability to generate two reads for one DNA fragment which is useful in order to determine the distance between two ends of the DNA fragment [31]. In fact, due to its low cost per base and high yield, Illumina becomes the most widely used and popular NGS platform. The output of Illumina sequencing is the highest among all NGS, making it suitable for multiplexing hundreds of samples at the same time [32].

2.2.1.3 Applied biosystems (AB) SOLiD sequencer

AB SOLiD refers to sequencing by oligonucleotide ligation and detection. It has been developed by Applied Biosystems (Life Technology) and became commercially available in 2007. The AB SOLiD sequencing approach differs from the other

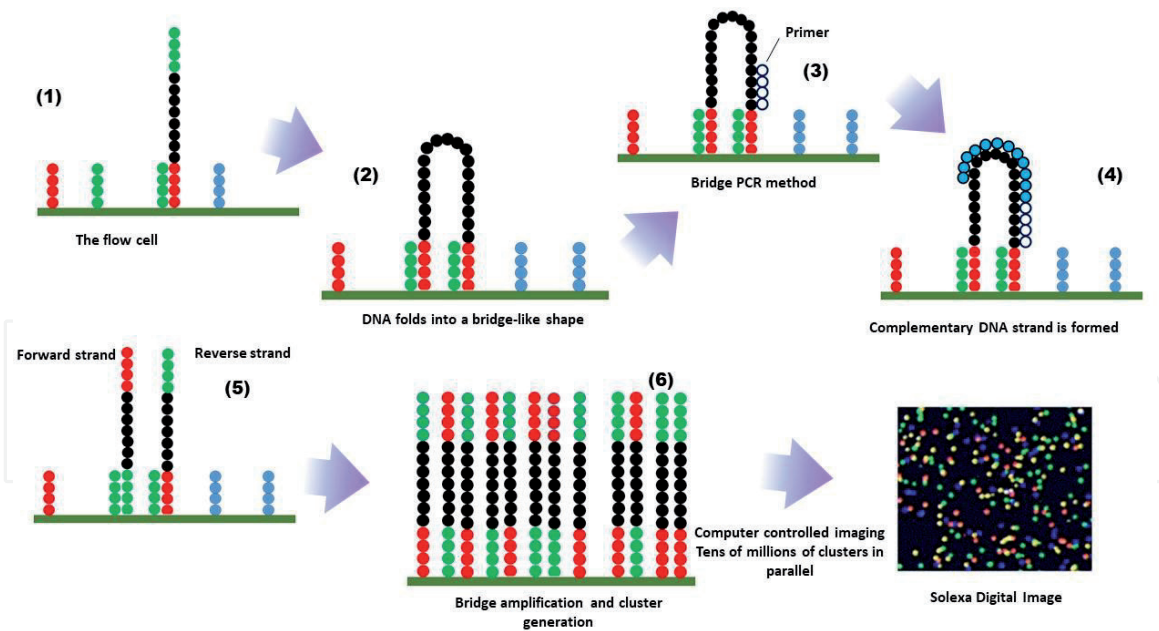


Figure 3. *Illumina/Solexa sequencing approach.* (1) The DNA templates with the attached adapter sequences are connected via a glass surface coated with oligos complementary sequences (2, 3, 4). DNA molecules fold over into a bridge shape and bridge PCR amplification is applied. (5) Bridge amplification and the formation of millions of copies or cluster formation. (6) Cluster sequencing is achieved through the process of cyclic reversible termination method. Finally, the resulting reads (tens of millions) are analyzed and the DNA sequence is recorded.

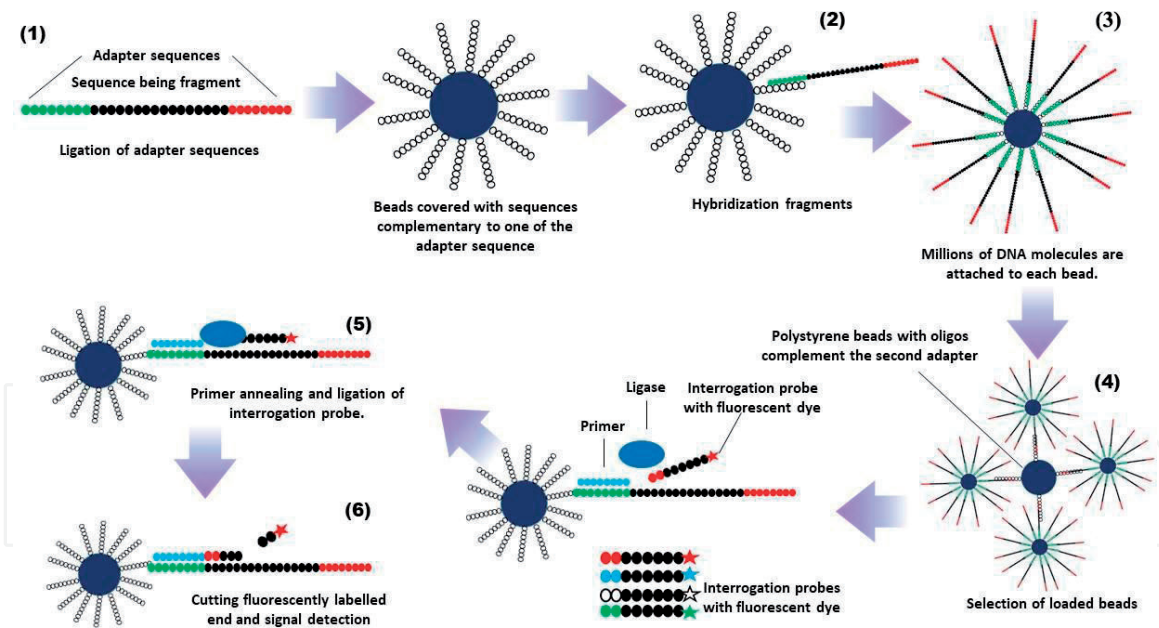


Figure 4. *Applied biosystems (AB) SOLiD sequencing approach.* (1) Preparation of DNA library from the sample and ligation of specific adaptor and the beads are then covered with the sequences complementary to one of the adapter sequences. (2) The adapter sequences will then bind to its complementary sequences on the beads. (3) The hybridization process resulted in the attachment of millions of DNA sequences to the bead. (4) Removal of the unloaded beads and selection of the loaded beads. (5) An interrogation probe contains six universal bases and two-base encoded probe. The universal bases are attached to the fluorescent label. (6) When an integrated probe is ligated with primers using DNA ligase, fluorescent light is generated and detected. This process is repeated several times till the targeted DNA is completely sequenced.

two next-generation sequencing technologies, Illumina, and 454 pyrosequencing. AB SOLiD platform relies on sequencing-by-oligo-ligation (SBL) (Figure 4), whereas others rely on sequencing-by-synthesis (SBS) [33]. In SOLiD sequencer,

the DNA library is prepared from the sample, and specific adaptor is then amplified by emPCR [34]. Instead of utilizing DNA polymerase, short nucleotides marked by DNA ligase known as interrogation probes are used. The interrogation probe contains six universal bases and two-base encoded probe. The universal bases are attached to the fluorescent label. When an integrated probe is ligated with primers using DNA ligase, fluorescent light is generated and detected. After the 5' end that is linked to the fluorescent label by cleavable linkage is cleaved and removed, thereby the next interrogation probe is connected. This process is repeated several times until the targeted DNA is completely sequenced. In fact, the read length of SOLiD is short about 85 bp leading to inaccurate read assembly as it requires more time for sequencing but it has the highest accuracy among other NGS [35]. Application of SOLiD includes whole genome sequencing, targeted sequencing, transcriptome, and epigenome [35].

2.2.1.4 Ion torrent sequencing

Ion Torrent has been launched in 2010 by Life Technology. Some authors have classified the Ion Torrent platform as a technique between the next-generation and the third-generation sequencing. This could be attributed to the dependence of this approach on optical sensors. However, it relies on chemical sensors that detect the hydrogen-ion concentration change that occurred during the incorporation of a nucleotide in the sequence [21]. Ion Torrent sequencing quality is high and stable due to the utilizing of a chemical sensor instead of fluorescence and camera. In addition, the Ion Torrent approach is characterized by its high speed and low cost compared with pyrosequencing and Illumina [35].

2.2.2 Third-generation sequencing

The major limitations of NGS are that the short-read length and the PCR bias are introduced by clonal amplification and the fluorescent-based signaling detection [21]. Therefore, the third-generation sequencing or single-molecule-sequencing technologies (SMS) overcome these limitations by dispensing PCR before sequencing, and the signal is captured in real time by monitoring the enzymatic reaction [36, 21]. The following sections discuss some TGS platforms.

2.2.2.1 Helicos biosciences (HeliScope)

The first single-molecule-sequencing (SMS) that has been introduced in 2008 is HeliScope. It is a fluorescent-based, single-molecule-sequencing platform. In HeliScope platform, the preparation step depends on preparing a single-strand DNA, and there is no need for PCR amplification in the preparation step. During sequencing, repetitive cycles of DNA polymerase and one labeled nucleotide are flowed, resulting in DNA template extension which depends on the flow of nucleotides. The labeled nucleotides are modified by attaching a poly-A tail in order to stop polymerase extension until the fluorescence that generates from the incorporated nucleotide is recorded by a CCD camera. Then unincorporated nucleotides are washed out and the fluorescent labels on the strand chemically removed, allowing for next base incorporation [37, 38]. HeliScope Genetic Analysis System platform allows the sequencing of RNA, and there is no need for converting them to cDNA. Furthermore, HeliScope Genetic Analysis System platform is in its infancy due to small read length (24–70 bases) and low data output (20 GB) [39].

2.2.2.2 PacBio technology/SMRT sequencer

Pacific Bioscience has launched a single-molecule real-time (SMRT) technology in 2010. It is a real-time, fluorescent-based, and single-molecule-sequencing platform. In SMRT, there is no need for PCR amplification during DNA preparation [36]. In this platform, a nanostructure known as zero-mode waveguide (ZMW) is utilized for real-time observation of DNA synthesis. During the sequencing process, a single-stranded template is used to synthesize the complementary. Unlike other NGS platforms, four different colored fluorescent labels are attached to the terminal phosphate group instead of attaching to a nucleotide, resulting in the release of a fluorescent signal during nucleotide incorporation [40]. Then the camera captures the fluorescent signal in real time (like a movie) [41]. In SMRT, the washing step between nucleotide flows is not required, resulting in increasing the nucleotide incorporation and improving the quality of sequencing [42]. SMRT has several advantages including fast sample preparation (hours instead of days like NGS), no need for PCR amplification during the preparation step, and longer-read length than any other next-generation sequencing platform [42].

2.2.2.3 Oxford Nanopore technology

Nanopore sequencing, developed by Oxford Nanopore Technology, relies on passing the DNA sequence through 1 nm diameter hole (nanopore) where electric current is applied. The electrical current of the pore is altered for each nucleotide, and signal is detected in real time [39]. Like other third-generation sequencing approaches, this technology does not require PCR amplification or chemical labeling of the sample [43]. In May 2015, Oxford Nanopore Technologies has introduced commercially the MinION. The MinION is a pocket-size portable, real-time detection of bases (fluorescent tag-free), has long-read length, and is a low-cost technology [44, 41, 45]. Interestingly, by utilizing this technology, samples can be sequenced in the field directly, instead of collecting samples and sequencing them in the lab, which means nanopore sequencing will make all other sequencing machines redundant [46, 44].

2.3 Metagenomic data analysis

Several bioinformatic tools were developed to analyze the metagenomic data at the molecular level (e.g., 16S rRNA), species level, and strain level. 16S rRNA sequence strategy is among the most common approaches to understand microbial taxonomy and phylogeny. This could be attributed to the stable functions of 16S rRNA gene over time, the existence of 16S rRNA in nearly all microorganisms, and its size which is enough for bioinformatics analysis [47, 48]. A number of bioinformatics tools are available for the analysis of 16S rRNA: QIIME, MOTHUR, DADA2, UPARSE, and minimum entropy decomposition (MED) [49]. The QIIME software is designed to analyze data generated on the Illumina or other NGS platforms via graphics and statistics. This involves the demultiplexing and quality filtering, OTU picking, taxonomic assignment, and phylogenetic reconstruction, and diversity analyses and visualizations [50, 51]. QIIME depends on the use of the PyCogent toolkit to identify misinterpretations and database deposition using raw sequencing results [51]. Operational taxonomic units (OTUs) can be generated from NGS data by UPARSE [52]. The UPARSE software acts by filtering and trimming reads into equal lengths, removing singleton reads and clustering the remaining reads [52].

Community sequence data can be analyzed by a flexible and comprehensive software package called MOTHUR. The MOTHUR package includes the following algorithms: DOTUR, SONS, TreeClimber, LIBSHUFF, D-LIBSHUFF, and UniFrac [50]. DADA2 is a suitable approach for correcting amplicon errors with no option to generate OTUs [53]. DADA2 uses a new quality-aware model of Illumina amplicon errors to improve the DADA algorithm [53]. MED is applied to solve the limitations of fine-scale resolution descriptions of microbial communities [54]. MED acts through partitioning the data set of amplicon sequences into homogenous OTUs for alpha- and beta-diversity analyses [54].

For species-level metagenomic data analysis, there are at least six metagenomic analysis software including MetaPhlAn2 [55], Kraken [56], CLARK [57], FOCUS [58], SUPERFOCUS [59], and MG-RAST [60]. All of these software programs can be used to profile organisms in metagenomic samples and to score their abundance. MetaPhlAn2 applies Bowtie2 and UCLUST [52, 61] as its main algorithms, whereas k-mers (DNA words of length k) is the core algorithm for Kraken and CLARK. On the other hand, FOCUS uses the NNLS (nonnegative least squares) to identify the microbial profile [49].

3. Conclusion

At the beginning, the metagenomic workflow was complicated that it requires many steps, sophisticated equipment, and qualified technicians to perform. Likewise, it was very expensive that not all scientists or laboratories were able to afford its cost. However, nowadays, due to the presence of many different competing companies and laboratories that led to the development of more efficient sequencing approaches, the metagenomic workflow became easier. It is easy now to study and identify organisms directly from their habitats without prior preparations. In terms of cost, NGS is also much cheaper, and with the appearance of third-generation sequencing approaches, it is not required to conduct sample sequencing. Surprisingly, sequencing can be carried out in the field by utilizing a pocket-size portable sequencer. The advancements in the field of metagenomics are amazing, and it became easier, cheaper, and faster.

Acknowledgements

This work was funded by the Deanship of Scientific Research at Princess Nourah Bint Abdulrahman University, through the Research Groups Program Grant no. RGP-1438-0004.

Conflict of interest

The authors declare that there are no conflicts of interest.

IntechOpen

Author details

Ahmed Shuikan¹, Sulaiman Ali Alharbi¹, Dalal Hussien M. Alkhalifah²
and Wael N. Hozzein^{3,4*}

1 Botany and Microbiology Department, College of Science, King Saud University, Riyadh, Saudi Arabia

2 Biology Department, Faculty of Science, Princess Nourah Bint Abdulrahman University, Riyadh, Saudi Arabia

3 Bioproducts Research Chair, Zoology Department, College of Science, King Saud University, Riyadh, Saudi Arabia

4 Botany and Microbiology Department, Faculty of Science, Beni-Suef University, Beni-Suef, Egypt

*Address all correspondence to: hozzein29@yahoo.com; whozzein@ksu.edu.sa

IntechOpen

© 2019 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. 

References

- [1] Xu J. Microbial ecology in the age of genomics and metagenomics: Concepts, tools, and recent advances. *Molecular Ecology*. 2006;**15**:1713-1731
- [2] Schlaberg R, Chiu CY, Miller S, Procop GW, Weinstock G. Microbiology Resource Committee of the College of American Pathologists. Professional Practice Committee and Committee on Laboratory Practices of the American Society for Microbiology, et al. Validation of metagenomic next generation sequencing tests for universal pathogen detection. *Archives of Pathology & Laboratory Medicine*. 2017;**141**:776-786
- [3] Escobar-Zepeda A, Vera-Ponce de León A, Sanchez-Flores A. The road to metagenomics: From microbiology to DNA sequencing technologies and bioinformatics. *Frontiers in Genetics*. 2015;**6**:348
- [4] Almeida OGG, DeMartinis OCP. Bioinformatics tools to assess metagenomic data for applied microbiology. *Applied Microbiology and Biotechnology*. 2019;**103**(1):69-82
- [5] Greninger AL. The challenge of diagnostic metagenomics. *Expert Review of Molecular Diagnostics*. 2018;**18**:605-615
- [6] Gu W, Miller S, Charles Y. Chiu. Clinical metagenomic next-generation sequencing for pathogen detection. *Annual Review of Pathology*. 2019;**14**:319-338
- [7] Greninger AL, Zerr DM, Qin X, Adler AL, Sampoleo R, Kuypers JM, et al. Rapid metagenomic next generation sequencing during an investigation of hospital-acquired human parainfluenza virus 3 infections. *Journal of Clinical Microbiology*. 2017;**55**:177-182
- [8] Simner PJ, Miller HB, Breitwieser FP, Monsalve GP, Pardo CA, Salzberg SL, et al. Development and optimization of metagenomic next-generation sequencing methods for cerebrospinal fluid diagnostics. *Journal of Clinical Microbiology*. 2018;**56**:e00472-e00418
- [9] Vincent AT, Derome N, Boyle B, Culley AI, Charette SJ. Next-generation sequencing (NGS) in the microbiological world: How to make the most of your money. *Journal of Microbiological Methods*. 2017;**138**:60-71
- [10] Quince C, Walker AW, Simpson JT, Loman NJ, Segata N. Shotgun metagenomics, from sampling to analysis. *Nature Biotechnology*. 2017;**35**(9):833
- [11] Young JM, Rawlence NJ, Weyrich LS, Cooper A. Limitations and recommendations for successful DNA extraction from forensic soil samples: A review. *Science & Justice*. 2014;**54**(3):238-244
- [12] Finley SJ, Lorenco N, Mulle J, Robertson BK, Javan GT. Assessment of microbial DNA extraction methods of cadaver soil samples for criminal investigations. *Australian Journal of Forensic Sciences*. 2016;**48**(3):265-272
- [13] Lim NY, Roco CA, Frostegård Å. Transparent DNA/RNA co-extraction workflow protocol suitable for inhibitor-rich environmental samples that focuses on complete DNA removal for transcriptomic analyses. *Frontiers in Microbiology*. 2016;**7**:1588
- [14] Gupta P, Manjula A, Rajendhran J, Gunasekaran P, Vakhlu J. Comparison of metagenomic DNA extraction methods for soil sediments of high elevation Puga hot spring in Ladakh, India to explore bacterial diversity. *Geomicrobiology Journal*. 2017;**34**(4):289-299
- [15] Mazziotti M, Henry S, Laval-Gilly P, Bonnefoy A, Falla J. Comparison of two

- bacterial DNA extraction methods from non-polluted and polluted soils. *Folia Microbiologica*. 2018;**63**(1):85-92
- [16] Van Djick EL, Auger H, Jaszczyszyn Y, Thermes C. Ten years of next-generation sequencing technology. *Trends in Genetics*. 2014;**30**:418-426
- [17] Head SR, Komori HK, LaMere SA, Whisenant T, Van Nieuwerburgh F, Salomon DR, et al. Library construction for next generation sequencing: Overviews and challenges. *Biotech*. 2014;**56**:61-68
- [18] Simon C, Daniel R. Construction of small-insert and large-insert metagenomic libraries. In: *Metagenomics*. New York, NY: Humana Press; 2017. pp. 1-12
- [19] Sanger F, Coulson AR. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *Journal of Molecular Biology*. 1975;**94**:441-448
- [20] Maxam AM, Gilbert W. A new method for sequencing DNA. *Proceedings of the National Academy of Sciences*. 1977;**74**:560-564
- [21] Schadt EE, Truner S, Kasarskis A. A window into third-generation sequencing. *Human Molecular Genetics*. 2010;**19**:R227-R240
- [22] Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences*. 1977;**74**:5463-5467
- [23] Hert DG, Fredlake CP, Barron AE. Advantages and limitations of next-generation sequencing technologies: A comparison of electrophoresis and non-electrophoresis methods. *Electrophoresis*. 2008;**29**:4618-4626
- [24] Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, et al. The sequence of the human genome. *Science*. 2001;**291**:1304-1351
- [25] Metzker ML. Sequencing technologies-the next generation. *Nature Reviews. Genetics*. 2010;**11**:31-46
- [26] Varshney RK, Nayak SN, May GD, Jackson SA. Next-generation sequencing technologies and their implications for crop genetics and breeding. *Trends in Biotechnology*. 2009;**27**:522-530
- [27] Sogin ML, Morrison HG, Huber JA, Mark Welch D, Huse SM, Neal PR, et al. Microbial diversity in the deep sea and the underexplored 'rare biosphere'. *Proceedings of the National Academy of Sciences*. 2006;**103**:12115-12120
- [28] Mardis ER. Next-generation DNA sequencing methods. *Annual Review of Genomics and Human Genetics*. 2008;**9**:387-402
- [29] Margulies M, Egholm M, Altman W, et al. Genome sequencing in microfabricated high-density picolitre reactors. *Nature*. 2005;**437**:376-380
- [30] Huse SM, Huber JA, Morrison HG, Sogin ML, Welch DM. Accuracy and quality of massively parallel DNA pyrosequencing. *Genome Biology*. 2007;**8**(7):R143
- [31] Fullwood MJ, Wei CL, Liu ET, Ruan Y. Next-generation DNA sequencing of paired-end tags (PET) for transcriptome and genome analyses. *Genome Research*. 2009;**19**(4):521-532
- [32] Glenn TC. Field guide to next generation DNA sequencers. *Molecular Ecology Resources*. 2011;**11**(5):759-769. Available from: <https://www.molecularecologist.com/next-gen-fieldguide-2016/>
- [33] McKernan K, Blanchard A, Kotler L, Costa G. Reagents, methods

- and libraries for bead-based sequencing. US20080003571; 2008
- [34] Shao K, Ding W, Wang F, Li H, Ma D, Wang H. Emulsion PCR: A high efficient way of PCR amplification of random DNA libraries in aptamer selection. *PLoS One*. 2011;**6**(9):e24910. DOI: 10.1371/journal.pone.0024910
- [35] Liu L, Li Y-H, Li S, Hu N, He Y, Pong R, et al. Comparison of next-generation sequencing systems. *Journal of Biomedicine & Biotechnology*. 2012;**2012**:251364. DOI: 10.1155/2012/251364
- [36] Korlach J, Bjornson KB, Chaudhuri BP, Cicero RL, Flusberg BA, Gray JJ, et al. Real-time DNA sequencing from single polymerase molecules. *Methods in Enzymology*. 2010;**472**:431-455
- [37] Harris T, Buzby P, Babcock H, et al. Single molecule DNA sequencing of a viral genome. *Science*. 2008;**320**:106-109
- [38] Zhang J, Chiodini R, Badr A, Zhang G. The impact of next-generation sequencing on genomics. *Journal of Genetics and Genomics*. 2011;**38**:95-109
- [39] Hart C, Lipson D, Ozsolak F, Raz T, Steinmann K, Thompson J, et al. Single molecule sequencing: Sequence method to enable accurate quantitation. *Methods in Enzymology*. 2010;**472**:407-430
- [40] Flusberg BA, Webster DR, Lee JH, Travers KJ, Olivares EC, Clark TA, et al. Direct detection of DNA methylation during single-molecule, real-time sequencing. *Nature Methods*. 2010;**7**:461-465
- [41] Timp W, Mirsaidov UM, Wang D, Comer J, Aksimentiev A, Timp G. Nanopore sequencing: Electrical measurements of the code of life. *IEEE Transactions on Nanotechnology*. 2010;**9**(3):281-294
- [42] Zhou X, Ren L, Li Y, Zhang M, Yu Y, Yu J. The next-generation sequencing technology: A technology review and future perspective. *Science China. Life Sciences*. 2010;**53**:44-57
- [43] Shi Y, Tyson GW, Eppley JM, DeLong EF. Integrated metatranscriptomic and metagenomic analyses of stratified microbial assemblages in the open ocean. *The ISME Journal*. 2011;**5**:999-1013
- [44] Hayden EC. Nanopore genome sequencer makes its debut. *Nature*. 2012;**10051**. DOI: 10.1038/nature.2012.10051
- [45] Laver T, Harrison J, O'Neill PA, Moore K, Farbos A, Paszkiewicz K, et al. Assessing the performance of the Oxford Nanopore technologies MinION. *Biomolecular Detection and Quantification*. 2015;**3**:1-8
- [46] Jain M, Fiddes IT, Miga KH, Olsen HE, Paten B, Akesen M. Improved data analysis for the MinION nanopore sequencer. *Nature Methods*. 2015;**12**:351-356
- [47] Patel JB. 16S rRNA gene sequencing for bacterial pathogen identification in the clinical laboratory. *Molecular Diagnosis*. 2001;**6**:313-321
- [48] Janda JM, Abbott SL. 16S rRNA gene sequencing for bacterial identification in the diagnostic laboratory: Pluses, perils, and pitfalls. *Journal of Clinical Microbiology*. 2007;**45**:2761-2764
- [49] Niu SY, Yang J, McDermaid A, Zhao J, Kang Y, Ma Q. Bioinformatics tools for quantitative and functional metagenome and metatranscriptome data analysis in microbes. *Briefings in Bioinformatics*. 2018;**19**(6):1415-1429. DOI: 10.1093/bib/bbx051
- [50] Schloss PD, Westcott SL, Ryabin T, et al. Introducing MOTHUR: Open-source, platform-independent,

community supported software for describing and comparing microbial communities. *Applied and Environmental Microbiology*. 2009;**75**:7537-7541

[51] Caporaso JG, Kuczynski J, Stombaugh J, et al. QIIME allows analysis of high-throughput community sequencing data. *Nature Methods*. 2010;**7**:335-336

[52] Edgar RC. UPARSE: Highly accurate OTU sequences from microbial amplicon reads. *Nature Methods*. 2013;**10**:996-998

[53] Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. DADA2: High resolution sample inference from Illumina amplicon data. *Nature Methods*. 2016;**13**:581-583

[54] Eren AM, Morrison HG, Lescault PJ, Reveillaud J, Vineis JH, Sogin ML. Minimum entropy decomposition: Unsupervised oligo typing for sensitive partitioning of high-throughput marker gene sequences. *The ISME Journal*. 2015;**9**:968-979

[55] Truong DT, Franzosa EA, Tickle TL, Scholz M, Weingart G, Pasolli E, et al. MetaPhlan2 for enhanced metagenomic taxonomic profiling. *Nature Methods*. 2015;**12**:902-903

[56] Wood DE, Salzberg SL. Kraken: Ultrafast metagenomic sequence classification using exact alignments. *Genome Biology*. 2014;**15**:R46

[57] Ounit R, Wanamaker S, Close TJ, Lonardi S. CLARK: Fast and accurate classification of metagenomic and genomic sequences using discriminative k-mers. *BMC Genomics*. 2015;**16**:236

[58] Silva GG, Cuevas DA, Dutilh BE, Edwards RA. FOCUS: An alignment free model to identify organisms in metagenomes using non-negative

least squares. *Peer J*. 2014;**2**:e425. DOI: 10.7717/peerj.425

[59] Silva GG, Green KT, Dutilh BE, Edwards RA. SUPER-FOCUS: A tool for agile functional analysis of shotgun metagenomic data. *Bioinformatics*. 2016;**32**:354-361

[60] Meyer F, Paarmann D, D'Souza M, Olson R, Glass EM, Kubal M, et al. The metagenomics RAST server-a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics*. 2008;**9**:386

[61] Langmead B, Salzberg SL. Fast gapped-read alignment with bowtie 2. *Nature Methods*. 2012;**9**:357-359