

# New generation genome sequencing methods

A. Ilgın Kekeç<sup>1</sup>, Nisa Sipahi<sup>2</sup>, Serkan İkiz<sup>1</sup>

<sup>1</sup> Istanbul University, Cerrahpaşa, Faculty of Veterinary Medicine, Department of Microbiology, Istanbul, Turkey  
<sup>2</sup> Düzce University Traditional and Complementary Medicine Application and Research Center, Düzce, Turkey

**ORCID ID of the author(s)**

AIK: 0000-0002-0821-8376  
NS: 0000-0001-8915-3545  
SI: 0000-0001-6502-0780

## Abstract

Sequencing procedures, which present information related to events in the living metabolism, are performed by genetic methods. Gene structure and genetic control mechanisms can be determined by sequencing methods, as they represent significant contributions to transcriptomic, ecological, and epidemiological studies. Thus, chromosomal abnormalities in living things with a whole genome analysis of different organisms (e.g., plants, bacteria, yeast, fungi, and viruses) are detected quickly, so that clinical disease can be diagnosed with reliable data from biomedical research. The nucleic acid sequencing processes have a history of about fifty years. With the technology developed in recent years, second and third-generation methods, known as next-generation sequencing (NGS), found their rightful place in the sector approximately 15 years ago. New-generation sequencing promotes inexpensive, routine, and comprehensive analyses of the living genome. In general, sequencing methods are separated as first-, second-, and third-generation sequencing methods, with the latter called 'new-generation sequencing;' our review focuses on these specific sequencing methods.

**Keywords:** DNA, Genome, New-generation sequencing, Sequencing methods

## Introduction

Nucleic acids are the main building blocks controlling metabolic events, heredity, and disease in living cells, while nucleic acids consist of simpler structures called nucleotides; the arrangement that comprise them is known as sequencing [1]: this contains important information about hereditary properties. In addition, determining nucleotide sequencing is necessary for biological and genetic research [2].

With the development of technology in recent years, sequencing and new-generation methods provide convenience in scientific research, medicine, and health fields. Although DNA and RNA sequencing date back nearly fifty years, large-scale mass parallel sequencing or next-generation sequencing have been studied for about 15 years [3]. NGS allows for a comprehensive analysis of the living genome, which is both inexpensive and widespread [4].

### Corresponding Author

A. Ilgın KEKEÇ  
Istanbul University-Cerrahpaşa, Faculty of Veterinary Medicine, Department of Microbiology, Büyükcçekmece, Istanbul, Turkey  
E-mail: ilginkekec@iuc.edu.tr

### Conflict of Interest

No conflict of interest was declared by the authors.

### Financial Disclosure

The authors declared that this study has received no financial support.

### Published

2022 April 10

Copyright © 2022 The Author(s)

Published by JOSAM

This is an open access article distributed under the terms of the Creative Commons Attribution-Non Commercial-NoDerivatives License 4.0 (CC BY-NC-ND 4.0) where it is permissible to download, share, remix, transform, and build upon the work provided it is properly cited. The work cannot be used commercially without permission from the journal.



Sequencing research emerged with an RNA molecule, as researchers looked for new methods for DNA to be sequenced as well [1]. The best-known method since the first appearance of sequencing is the Sanger method. In 1977, Frederick Sanger developed a technology based on chain termination. Thus, the first human genome project was carried out with this method [5], also known as Dideoxy or chain termination reactions. In the same year, Maxam and Gilbert developed another method based on chemical modification of DNA [3,6], known as the multidirectional sequence: DNA is cut from specific base sequences using chemicals and different length fragments and is then 'harvested.' In this method, forty clones can be analyzed from one sequencing gel. Nevertheless, it is the less preferred method than that of Sanger, since dangerous chemicals may be used. The Gilbert and Maxam-Sanger sequencing methods are defined as first-generation sequencing: they have been used in successful research involving human, animal, and plant genome work for the past 20 years. However, they are laborious, which led to next-generation sequencing methods; the ideal DNA sequencing technology must be accurate, easy, and inexpensive [5]. The genetic technology of next-generation sequencing is based on simultaneous processing of each part of a DNA molecule from a single living cell (consisting of millions of harmonious fragments) [7]. The technology has become an indispensable part of biological research in many fields with obvious advantages, such as obtaining data in a short time, which are richer and more reliable [1, 4]. At present, a variety of new-generation sequencing technologies are available, creating a library with many DNA fragments by segmenting the selected target area of genomic DNA from the biological material to be assessed via enzymes. Reproduction of the DNA fragments makes up the library, with an analysis of raw data by sequenced replicated fragments [8]. The next-generation methods are generally divided into second- and third-generation [9,10]. Roche 454 Genome Analyzer, Illumina, Inc, San Diego, CA, USA, SOLID, and IonTorrent (Thermo Fisher Scientific, Waltham, MA, USA) use second-generation sequencing for the formation of mRNA, a small RNA profile, with a genome length characterization, chromatin structure, ancestral DNA microbiology, and similar molecular characterization, as well as microbial communities [11].

#### **Emulsion PCR**

Mold DNA amplification in next-generation sequencing is performed with emulsion PCR, a template amplification cation in multiple platforms. The basic principle of EmPCR is dilution and division of template molecules in water-oil emulsion and droplets [11]. EmPCR allows for the amplification of DNA molecules in droplets of physically separated picoliter volumetric water, preventing formation of inefficient chimera and other artifacts in DNA sequences. Droplets are used in analytical applications, e.g., number determination with digital PCR and preparative applications: molecular evolution and genome-scale DNA [12].

#### **Second-Generation Sequencing Methods**

Given that time is valuable in diagnostic research, it was inevitable to develop the first sequencing technologies. SGSs on the market were more advantageous than first-generation technologies, a result of millions of parallel short readings,

accelerating the sequencing process, a low sequencing cost, and direct detection without any need for electrophoresis [9].

#### **Roche 454 Sequencing Technology**

The importance of microbial communities for human and environmental health led researchers to develop methods for these communities. The most common and cost-effective is sequencing targeted genetic elements. The amplicon sequence of taxonomic marker genes, such as the 16s rRNA gene in bacteria contains fungi and the 18S rRNA gene in eukaryotes. This provides a community population count: researching functional diversity is accomplished through targeting functional genes: one method is the Roche 454 sequencing technology [13].

The technology was first patented by Melamede in 1989 [10, 14]. Moreover, 454 Life Sciences, the first NGS technology on the market, became Roche 454 (Branford, CT, USA) based on the pyrosequencing method, or that of detecting pyrophosphate released from the new synthetic DNA chain (formed after each nucleotide combination) [9, 10].

Although this method of pyrosequencing has some disadvantages, it has been accepted in metagenomic studies on complex microbial communities. Among its disadvantages are high cost, errors in reading complex sequences, inefficiency in loading limitations due to bead-based DNA molecules, and a limited number of reads [15,16].

#### **Illumina Genome Analyzer**

The Illumina NGS technology, first introduced by Solexa and later purchased by Illumina, is a system based on the use of synthesis with reversible stain terminators [9, 10].

As with the 454 system, adapters are placed at the ends of the DNA sequence. However, the prior system uses beads and EmPCR, while Illumina uses planar solid glass support. During the first step of Illumina technology, DNA samples are broken down into random sequences, and adapters are connected to both ends of each array. These adapters connect to the corresponding complementary sequence. In the second stage, each sequence fixed to the solid plate is strengthened with PCR bridge amplification. A series of sequences formed from the same original is called a cluster, and each cluster contains approximately a million copies of the same original series. In the final stage, Illumina covers the synthesis approach, using reversible terminators to determine each nucleotide in the sequence [9, 10, 17].

Illumina is a commonly used NGS method worldwide [18]. Its technology evolved rapidly, providing high efficiency at low cost with improved read lengths and true paired end reads. One drawback is that it has extra requirements for sample load control, with a surfeit in these clusters that overlap and create incorrect sequencing. Overall, the error rate is about 1% [13, 18].

#### **Ion Torrent**

With the commercialization of the Ion Torrent Personal Genome Machine by Life Technologies (Carlsbad, CA, USA), this new-generation sequencing method simulates 454 pyrosequencing technology. However, Ion Torrent does not use fluorescent-labeled nucleotides as other technologies do; it is based on the hydrogen ion released during sequencing [9]. Accordingly, when new nucleotides are attached after pyrophosphate division, a proton is released. The well plate ensures that proton release is localized and maintained. The

signal is proportional to the number of free protons, enabling the sequence of homopolymeric regions of template DNA. Data collection uses a complementary semiconductor metal oxide sensor chip (CMOS), which can measure millions of sequence reactions at once [10].

The most important aspect of the Ion Torrent Personal Genome Machine (PGM), Ohio University, Athens, OH, USA, is that it uses synthesis sequencing strategy as an alternative approach, is low-cost, high-speed technology, with a medium-length read time [19-21].

### **SOLiD**

SOLiD (Supported Oligonucleotide Ligation and Detection, Carlsbad, CA, USA) is a technology based on ligation biochemistry, and on the hybridization of 16 different props, the first two nucleotides which are known and capable of creating four different radiations on the template DNA chain, given the first two nucleotides, and the connection of probes to each other (with ligation) [5]. This system was introduced in 2006 by Agencourt, which later changed to Applied Biosystems, Bedford, MA, USA. The DNA fragments to be sequenced are connected to adaptor molecules and attached to the beads. Cloning is carried out with EmPCR [10, 22].

The long read times in SOLiD chemistry are a distinct disadvantage of this technology, with short read times and inaccuracy. However, in the latest releases, this has been improved [22].

### **Third-Generation Sequencing Methods**

Second-generation sequencing technologies are often used compared to first-generation sequencing. Yet, second-generation technologies may require PCR, so they become disadvantageous in time and cost [10]. As such, third-generation sequencing technologies (TGSs) were developed. Long-range DNA sequencing and mapping technologies created a new era in high-quality genome sequencing. Unlike second-generation sequencing, short readings with a length of several hundred base pairs, third-generation single-molecule technologies produce more than 10,000 bp of reading and map more than 100,000 bp of molecules [23, 24].

### **Helicos Genetic Analysis Systems**

The commercial application of the single-molecule sequencing method, first introduced by Helicos Genetic Analysis Systems, Cambridge, MA, USA, made it possible to sequence DNA directly without creating DNA amplification [25]. Although it has similarities to Illumina technology in its fluorescence detection method, there is no need for the DNA amplification stage [2]. It is based on SBS, unlike many second-generation methods based on cloning [26].

The Helicos library is easier to prepare than sequencing technologies, as preparation for ligation and amplification is not required to form the library. Helicos systems are commercial TGSs based on single-molecule fluorescent sequencing: it determines quantities of RNA molecules without converting them to cDNA [22, 27].

### **Pacific BioSciences Single-Molecule Real-Time Sequencing**

Single-molecule real-time (SMRT) sequencing was first developed by Pacific Biosciences, Menlo Park, CA, USA, based

on real-time monitoring of modified enzymes, unlike the Helicos single-molecule sequencing method [6].

SMRT sequencing is now widespread in basic sciences and in applied realms, with field research involving agriculture, environment, and medicine. SMRT sequencing offers significant advantages over current short-read DNA technologies and for sequencing non-oxidized DNA molecules [28]. Its advantages are that it facilitates the sequencing of long DNA molecules with high accuracy. The SMRT sequence is a new resolution-level technology that examines the molecular mechanisms of living cells [29]. SMRT sequencing being error-prone is incorrect, as many studies show evidence that it is diagnostically useful, with advantages over short-read sequencers. The SMRT sequencing technology was applied in breast cancer cell models to identify new gene fusion events, with known oncogenes. It has been used in research, and will be used in the diagnosis of related diseases in the future [26].

### **Nanopore Sequencing**

The sequencing method of Oxford Nanopore Technologies™ (ONT) Oxford, UK determines the order of nucleotides in a DNA sequence. Launched in 2014 by Oxford Nanopore Technologies MinION, the device produces longer readings with better performance [9]. The device is four inches long and connects to a USB 3.0 port of any PC [30]. Protein channels with nano-sized diameter dimensions (nanopores) are embedded in the lipid barrier of the sequencing system. Their biological task is to facilitate ion transitions, as nanopore sequencing is based on measuring voltage generated by these ions, as they pass through the pores [2, 6].

In this sequencing technology, the first strand of a DNA molecule is connected to the complementary strand via a molecule. The DNA fragment goes through a protein nanopore, producing a variation of ionic current, based on differences in moving nucleotides; this occupies pores when DNA is transformed through them. The change of the ionic current is gradually recorded in the graphic model and interpreted to describe the sequence [31, 32].

Nanopore sequencing is advantageous, as it is a low-cost and small-size device; data are displayed on screen and created without completing the work. The lysate can be sequenced without PCR amplification or a chemical labeling step. This minimized workforce cost in sequencing technologies until recently. However, the % error rate must also be improved [33].

### **Conclusion**

Since the invention of the sequencing technologies, many developments were made in advancing science and technology. The Maxam and Gilbert method was behind the Sanger method, in terms of use of dangerous chemicals: the Sanger sequencing method has become a movement over time. Second-generation sequencing was created to reduce some of the cost-, labor-, time-related losses, and obtaining more data in a shorter time. Second-generation sequencing methods provide convenience in diagnosis and scientific research. The short read time of the SGS technologies, the need for bioinformatics equipment, and the clonal amplification processes of PCR has led to development of these technologies and third-generation sequencing; they have found their place in scientific research.

Instead of cloned sequencing, the TGS technologies sequence a single template DNA molecule and yield faster results in a shorter time, using fewer biochemicals. In this sense, it is more economical.

The NGS technologies are now the starting point for some research areas, as well as studying and analyzing biological sequences. However, they must be improved, in terms of reducing the margin of error, and as such, used more frequently.

## References

- Kizmaz MZ, Paylan İC, and Erkan S. Historical development of DNA sequencing. *Gaziosmanpaşa University Gbad*. 2017;6(2):47-53.
- Darcan C, Türkyılmaz O. Overview of next generation sequencing technology. *Bilecik Şeyh Edebali University J of Sci*. 2018;5(1):41-9.
- McCombie WR, McPherson JD, Mardis ER. Next generation sequencing technologies. *Cold Spring Harb Perspect Med*. 2019;9(11):a036798.
- Ustek D, Abaci N, Sirma S, Cakiris A. New generation DNA sequencing. *Experimed*. 2011;1(1):11-8.
- Akyol İ, Yildiz MA, Tutar E. Yeni nesil nükleotid dizileme metodlarının biyokimyasal temelleri. *Tarım ve Doga Dergisi*. 2017;20(1):1-15.
- Liu L, Li Y, Li S, Hu N, He Y, Pong R, et al. Comparison of next-generation sequencing systems. *J Biomed Biotechnol*. 2012;2012:1-11. doi: 10.1155/2012/251364.
- Kamps R, Brandão RD, Bosh BJ van den, Paulussen ADC, Xanthoulea S, Blok MJ, et al. Next-Generation sequencing in oncology: Genetic diagnosis, risk prediction and cancer classification. *Int J Mol Sci*. 2017;18(2):308. doi: 10.3390/ijms18020308.
- Dogan M, Eroz R, Yuce H, Ozmerdivenli R. The known about next-generation sequencing (NGS). *Duzce Med J*. 2017;19(1):27-30.
- Kchouk M, Gibrat J, Elloumi M. Generations of sequencing technologies: from first to next generation. *Biol Med*. 2017;9(3):1-8. doi: 10.4172/0974-8369.1000395
- Brujijns B, Tiggelaar R, Gardeniers H. Massively parallel sequencing techniques for forensics: a review. *Electrophoresis*. 2018;39(21):2642-54.
- Kanagal-Shamanna R. Emulsion PCR. *Techniques and Applications. Clinical applications of PCR*. Humana Press, New York, NY. 2016;1392:33-42.
- Verma V, Gupta A, Chaudhary VK. Emulsion PCR made easy. *BioTechniques*. 2020;69(1):64-9.
- Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. DADA2: high-resolution sample inference from Illumina amplicon data. *Nat Methods*. 2016;13(7):581-3.
- Melamede RJ. U.S. Patent No. 4,863,849. Washington, DC: U.S. Patent and Trademark Office. 1989
- Luo C, Tsementzi D, Kyripides N, Read T, Konstantinidis KT. Direct comparisons of Illumina vs. Roche 454 sequencing technologies on the same microbial community DNA sample. *PLoS One*. 2012;7(2):1-12.
- White RA, Callister SJ, Moore RJ, Baker ES, Jansson JK. The past, present and future of microbiome analyses. *Nat Protocols*. 2016;11(11):2049-53.
- Manley LJ, Ma D, Levine SS. Monitoring error rates in Illumina sequencing. *Journal of biomolecular techniques: JBT*. 2016;27(4):125.
- Schirmer M, D'Amore R, Ijaz UZ, Hall N, Quince C. Illumina error profiles: resolving fine-scale variation in metagenomic sequencing data. *BMC Bioinformatics*. 2016;17(1):1-15.
- Song L, Huang W, Kang J, Huang Y, Ren H, Ding K. Comparison of error correction algorithms for Ion Torrent PGM data: application to hepatitis B virus. *Sci Rep*. 2017;7(1):1-11.
- Adamiak J, Otlewska A, Tafer H, Lopandic K, Gutarowska B, Sterflinger K, Piñar G. First evaluation of the microbiome of built cultural heritage by using the Ion Torrent next generation sequencing platform. *Int Biodeter Biodegradation*. 2018;131:11-8.
- Guerrero-Sanchez VM, Maldonado-Alconada AM, Amil-Ruiz F, Verardi A, Jorrin-Novo JV, Rey MD. Ion Torrent and Illumina, two complementary RNA-seq platforms for constructing the holm oak (*Quercus ilex*) transcriptome. *PLoS One*. 2019;14(1):1-18.
- Ambardar S, Gupta R, Trakroo D, Lal R, Vakhlu J. High throughput sequencing: an overview of sequencing chemistry. *Indian J Microbiol*. 2016;56(4):394-404.
- Hayden EC. Nanopore genome sequencer makes its debut. *Nat News*. 2012. doi: 10.1038/nature.2012.10051.
- Lee H, Gurtowski J, Yoo S, Nattestad M, Marcus S, Goodwin S, et al. Third-generation sequencing and the future of genomics. *BioRxiv*. 2016;1-20. doi: 10.1101/048603.
- Thompson JF, Steinmann KE. Single molecule sequencing with a HeliScope genetic analysis system. *Curr Protoc Mol Biol*. 2010;92(1):7-10.
- Ardui S, Ameer A, Vermeesch JR, Hestand MS. Single Molecule Real-Time (SMRT) Sequencing Comes of Age: Applications and utilities for medical diagnostics. *Nucleic Acids Res*. 2018;46(5):2159-68.
- Hart C, Lipson D, Ozsolak F, Raz T, Steinmann K, Thompson J, Milos PM. Single molecule sequencing: sequence method to enable accurate quantitation. *Methods Enzymol*. 2010;472:407-30. doi: 10.1016/S0076-6879(10)72002-4.
- Hestand MS, Ameer A. The versatility of SMRT sequencing. *Genes*. 2019;10(1):24. doi: 10.3390/genes10010024.
- Ameer A, Che H, Martin M, Bunikis I, Dahlberg J, Hoijer I, et al. De novo assembly of two Swedish genomes reveals missing segments from the human GRCh38 reference and improves variant calling of population-scale sequencing data. *Genes*. 2018;9(10):486. doi: 10.3390/genes9100486
- Laehnemann D, Borkhardt A, McHardy AC. Denoising DNA deep sequencing data high-throughput sequencing errors and their correction. *Brief Bioinformatics*. 2015;17(1):154-79.
- Jain M, Hugh EO, Paten B, Akeson M. The oxford nanopore MinION: Delivery of nanopore sequencing to the genomics community. *Genome Biol*. 2016;17(1):1-11.
- Lu H, Giordano F, Ning Z. Oxford nanopore MinION sequencing and genome assembly. *Genomics Proteomics Bioinformatics*. 2016;14(1):265-79.
- Laver T, Harrison J, O'Neill PA, Moore K, Farbos A, Paszkiewicz K, Studholme DJ. Assessing the performance of the Oxford Nanopore Technologies MinION. *Biomol Detect Quantif*. 2015;3:1-8. doi: 10.1016/j.bdq.2015.02.001.

The National Library of Medicine (NLM) citation style guide has been used in this paper.