

Overviews of “next-generation sequencing”

Enas M Mostafa¹

Dalia M Sabri²

Sanaa M Aly^{1,2}

¹Department of Forensic Medicine and Clinical Toxicology, Faculty of Medicine, ²Biotechnology Research Center, Suez Canal University, Ismailia, Egypt

Abstract: DNA sequencing technology has a rich diverse history. Over the past several years, next-generation sequencing (NGS) has demonstrated its potential to accelerate research in the field of life sciences in general and forensic sciences in particular. Although the abilities of these new technologies allow cheap and comprehensive analyses within a short time, other important characteristics other than the technologies themselves (such as the quantity and quality of data) are generated. These characteristics present a set of challenges for experimental design, data analysis, and interpretation. Moreover, other improvements are needed to make NGS technology more attractive for routine forensics work. This review aims to shed light on the development and advancements of NGS from both historical and contemporary views. Advantages and limitations of NGS are mentioned and NGS in forensic sciences is summarized.

Keywords: next-generation sequencing, NGS, Sanger sequencing, second-generation, third-generation sequencing

Introduction

Nucleic acid sequencing is a tool that helps in determining the exact nucleotides order in DNA or RNA molecule.¹ Fifteen years have been passed between DNA double helix discovery (1953) and the first experiment based on DNA sequencing (1968). This gap was caused by multiple factors such as the similar chemical properties of different DNA molecules, making it appear difficult to separate them.²

The following period witnessed that DNA sequencing technologies and applications act as the engine of the genome era. This subsequently widens the amount of genome data and the range of research areas with many applications.³ In the early 1970s, the first DNA sequences were obtained using laborious methods such as two-dimensional chromatography.⁴

First-generation sequencing

Frederick Sanger developed DNA sequencing technology in 1977, which was based on the chain-termination or dye-termination method. It is known as Sanger sequencing. Another sequencing technology has been developed by Maxam and Gilbert based on the chemical DNA modification and cleavage at certain bases.⁵ Sanger sequencing was adopted as the primary technology in the “first generation” of commercial sequencing applications because of its high efficiency and low radioactivity. DNA sequencing was laborious at that time.^{1,3}

After years of improvement, Applied Biosystems (ABI) introduced the first automatic sequencing machine (namely AB370) in 1987, adopting capillary

Correspondence: Sanaa M Aly
Department of Forensic Medicine and
Clinical Toxicology, Faculty of Medicine,
Suez Canal University, Round Road,
Ismailia 41522, Egypt
Email sasydayem@hotmail.com

electrophoresis (CE) that made the sequencing faster and more accurate. Emerged in 1998, the automatic sequencing instruments and associated software using the capillary sequencing machines and Sanger sequencing technology became the main tools for the completion of Human Genome Project in 2001.³ Since the completion of that project, there is an increase in the demand for cheaper and faster sequencing methods. This demand has driven the development of next-generation sequencing (NGS).¹

Next-generation sequencing

NGS technology refers to non-Sanger-based high-throughput DNA sequencing technology. Millions or billions of DNA molecules can be sequenced in parallel, thereby increasing the throughput sequencing.⁶ Consequently, the entire genome could be sequenced in less than 1 day. Many NGS platforms have been developed with low-cost and high-throughput sequencing.¹ NGS technology includes second-generation sequencing technology as well as third-generation sequencing technology.⁶

Second-generation sequencing

These technologies provide a significant time saving because of highly streamlined sample preparation before sequencing, with a minimal need for associated equipment in comparison with the highly automated and multistep pipelines necessary for clone-based sequencing.

Each one of the following technologies aims to amplify single strands of a fragment library and make sequencing reactions on the amplified strands. The fragment libraries are obtained by annealing platform-specific linkers to blunt-ended fragments generated directly from DNA source of interest. In these technologies, no bacterial cloning step is required to amplify the genomic fragment in a bacterial intermediate as is done in conventional approaches. This is because the presence of adapter sequences that help the molecules to be selectively amplified by polymerase chain reaction (PCR).⁷

Roche 454 Genome Sequencer FLX System

In 2005, Roche introduced the 454 Genome Sequencing System, the world's first pyrosequencing-based high-throughput sequencing system.⁶

The 454 technology, the first to be released in the market, allows shotgun sequencing of whole genomes and avoids the cloning requirement by taking advantage of a highly efficient *in vitro* DNA amplification method known as emulsion PCR, where each DNA fragment-carrying streptavidin bead is

captured into separate emulsion droplets. The droplets act as individual amplification reactors, producing multiple copies of a unique DNA template per bead. Each bead is transferred into a picotiter plate well and the templates are analyzed using a pyrosequencing reaction, which is a sequencing-by-synthesis (SBS) technique that measures the release of inorganic pyrophosphate by chemiluminescence. The release of inorganic pyrophosphate is detectable by light produced by a chemiluminescent enzyme present in the reaction mix. The sequence of DNA template is determined from the order of integrated nucleotides that appeared in pyrogram.⁸

The most outstanding advantage of Roche is its speed: it takes only 10 hours from sequencing start till completion. The read length is also a distinguished character compared with other NGS system. But the high cost of reagents remains a challenge for Roche 454.³

ABI/SOLiD second-generation sequencing system

In 2007, ABI introduced the SOLiD second-generation sequencing system.⁶ A massively parallel sequencing by hybridization–ligation that begins with an emulsion PCR single-molecule amplification step similar to that used in the 454 technique. The amplification products are transferred onto a glass surface where sequencing occurs by sequential rounds of hybridization and ligation with 16 dinucleotide combinations labeled by four different fluorescent dyes. Using the four-dye encoding scheme, each position is effectively probed twice, and the identity of the nucleotide is determined by analyzing the color that results from two successive ligation reactions.⁸

However, SOLiD's computational infrastructure is expensive and not trivial to use; it needs an air-conditioned data center, computing cluster, skilled staff, distributed memory cluster, fast networks, and batch queue system.³

Illumina/Solexa sequencing technology

Illumina released Solexa sequencing technology that was based on cloning-free DNA amplification with attachment between single-stranded DNA fragments and a solid surface known as single-molecule DNA templates. The templates are sequenced using a DNA sequencing that employs reversible terminators with removable fluorescent moieties and special DNA polymerases that can incorporate these terminators into growing oligonucleotide chains. The template sequence of each cluster is deduced by reading off the color based on fluorescent-labeled terminators at each nucleotide.⁸

The Illumina approach produces shorter sequence reads than pyrosequencing. Moreover, substitution errors have been noted in its sequencing data due to the modified DNA polymerases and reversible terminators.⁸

Illumina/HiSeq 2000

Illumina launched HiSeq 2000 in early 2010, which adopts the same strategy as Illumina/Solexa sequencing technology. It is the cheapest in comparison with 454 and SOLiD and it could handle thousands of samples simultaneously.³

In conclusion, the longest read length is in Roche 454 system, the highest accuracy is in SOLiD system, and the biggest output and the lowest reagent cost are in the Illumina HiSeq 2000.³ More advanced technologies were produced later, eg, Ion Torrent and MiSeq instruments.

Ion Torrent/Ion Personal Genome Machine

At the end of 2010, Ion Personal Genome Machine was released by Ion Torrent, which was based on the semiconductor sequencing technology. A proton is released when a nucleotide is integrated into the DNA molecules by the polymerase. Based on pH changes, Personal Genome Machine recognized the addition of nucleotide. The chip was flooded with one nucleotide each time after another. If it is not the correct nucleotide, then no voltage will be found.³

Ion Torrent, a faster and low-cost sequencer based on the semiconductor technology, was introduced. This sequencer does not rely on fluorescence, chemiluminescence, or enzyme cascades for sequencing signal detection.⁶

Illumina/MiSeq

In 2011, MiSeq (a bench top sequencer) was launched by Illumina, which was based on the SBS technology. The functions of cluster generation, SBS, and data analysis have been integrated in a single instrument and can make complete analysis within a single day.³ It is a lower throughput fast-turnaround instrument aiming at smaller laboratories and the clinical diagnostic market.⁹ The most prominent parts of MiSeq are the highly integrated data and wider range of application.³

Third-generation sequencing

New insight in the sequencing has been occurred after the coming out of third-generation sequencing. It has two main criteria that save time. First, no need for PCR before sequencing so the time of DNA preparation is decreased. Second, the signals' monitoring is performed directly

during the enzymatic reaction of nucleotide additions; it is not important either it is fluorescent (Pacific Bioscience [PacBio]) or it is electric current (Nanopore).³

Pacific Bioscience

Single-molecule real time is the method developed by PacBio, which was based on modified enzyme and direct monitoring of the enzymatic reaction in real time.³

Millions of zero-mode waveguides constitute single-molecule real time cell, embedded with the set of DNA template and enzymes. Then, the camera in the machine will capture signal in a movie format. So that the fluorescent signal and the signal difference along time could be detected. These structural variances are useful in epigenetic studies.³

PacBio has several advantages in comparing with second generation. First, the sample preparation takes 4–6 hours instead of days. It also does not need PCR step in the preparation, which reduces bias and error caused by PCR. Second, the runs are finished within a day so the turnover rate is quite fast. Third, the read length is 1,300 bp in average that is longer than any second-generation sequencing technology. However, the throughput of the PacBio is lower than second generation.³

Nanopore sequencing

Nanopore sequencing is one of the third-generation sequencing methods. It involves the use of a tiny biopore with diameter in nanoscale found in protein channel embedded on lipid bilayer, which facilitates ion exchange.³ It also involves putting a thread of single-stranded DNA across α -hemolysin; a protein isolated from *Staphylococcus aureus*, undergoes self-assembly to form a heptameric transmembrane channel with extraordinary tolerance to high voltage. An ionic flow is applied continuously in this kind. Current disruption can be captured by a standard electrophysiological technique. Readout is relied on the size difference of deoxyribo-nucleoside-monophosphate. Thus, discrimination depends on characteristic current modulation.³

The most important advantages of this technology “nanopore sequencing” are minimal sample preparation, with no need to nucleotides, polymerases, or ligases, and the potential of very long read lengths. It will offer a tremendous cost reduction if there is a successful device. It makes inexpensive rapid DNA sequencing because of its unique analytical capability.¹⁰

Applications of NGS

It allows the rapid identification of causal mutations at single-nucleotide resolution even in complex cases.¹¹ It could provide

a deeper understanding of microbial basic biology, taxonomy, evolution, and their roles in environmental ecosystems and human health.¹²

It opened the door in front of new areas of biology, including the ancient genomes investigation, identification of unknown etiological agent, and ecological diversity characterization. It could help in genome resequencing studies either to characterize accurately strains or isolates or to give comprehensive view to the genome variation in clinical isolates of pathogenic microbes and viruses. It could help to detect genome-wide patterns of methylation “epigenomic variation” and how these patterns change through the course of an organism’s development, in the context of disease, and other influences.⁷

The most needed application of NGS is the ability to rapidly read out the results with potential of making a combination with other experiment results such as in correlative analyses of genome-wide methylation, histone binding patterns, and gene expression. The power in these analyses is to begin unlocking the secrets of the cell.⁷ More applications of NGS, more than those covered here, are yet to come.

Advantages of NGS

The large numbers of short reads produced by NGS offer many opportunities for the development of new applications that benefit from the particular data format. It is called sequence census applications. In this approach, sequence data are used to reveal sequence polymorphisms in the template, the abundance of reads is used as a quantitative measure and reveals the internal structure of the template, eg, the presence of exons and introns.⁸

The enormous reads generated by NGS enabled the sequencing of entire genomes at an unprecedented speed, and thus it has been widely used in various fields of life sciences.⁶

These technologies do not require bacterial cloning of DNA fragments; they rely on the preparation of NGS libraries in a cell-free system. Instead of hundreds of sequencing reactions, NGS can perform thousands-to-many millions of sequencing reaction at the same time.⁶

The cost has a great impact on implementation. Therefore, NGS provides cheaper and higher throughput alternative to traditional Sanger sequencing. It is just a day that is needed to make whole small genomes. Thus, that facilitates the genes’ and regulatory elements’ discovery which related to disease.^{1,13}

The conventional method (Sanger sequencing) will likely remain the technology of choice for the immediate future.

But very soon, large-scale projects will quickly come to depend entirely on NGS.¹³

Limitations of NGS

Although NGS is considered as a time and money-saving tool in comparison with conventional sequencing, it is still expensive for many laboratories. The advance in bioinformatics is urgently needed for accurate data analysis and avoiding sequence errors such as those in homopolymer regions (spans of repeating nucleotides) based on certain NGS platforms.¹

NGS in forensic sciences

Compared with other fields of life sciences, forensic DNA analysis is confronted with template of low copy number, highly degraded and contaminated samples. Thus, there is an urgent need for high accuracy and reproducibility with time and cost considerations. The main methods that implemented in contemporary forensic DNA analyses are PCR and CE-based fragment analysis methods to detect variation in short tandem repeat. However, CE-based analysis has its limitations such as low resolution of genotyping of current markers (such as mitochondrial DNA) and mixture analyses, the inability to analyze multiple genetic polymorphisms in a single reaction using a single workflow, and loss of useful genomic information from degraded DNA samples. These limitations prompt the forensic scientists worldwide to explore the usefulness of NGS technology for forensic studies and practical applications.⁶

Robust validation processes are mandatory before any new procedure will be introduced to any accredited forensic laboratory. In addition, selecting the right sequencing platform, that meets the demands of sample size, coverage, cost, and accuracy, is also a fundamental need for forensic DNA applications. Therefore, huge numbers of studies assessed these new technologies. Moreover, the research in this interesting area will continue to advance.¹⁴

Disclosure

The authors report no conflicts of interest in this work.

References

1. Grada A, Weinbrecht K. Next-generation sequencing: methodology and application. *J Invest Dermatol*. 2013;133(8):e11.
2. Hutchison CA. DNA sequencing: bench to bedside and beyond. *Nucleic Acids Res*. 2007;35(18):6227–6237.
3. Liu L, Li Y, Li S, et al. Comparison of next-generation sequencing systems. *J Biomed Biotechnol*. 2012;2012:1–11. doi:10.1155/2012/251364.
4. Zhang X, Tan J, Yang M, et al. The date palm genome project in the Kingdom of Saudi Arabia. In: Manickavasagan A, Essa MM, Sukumar E, editors. *Dates: Production, Processing, Food, and Medicinal Values*. Boca Raton: CRC Press; 2012:29–44.

5. França LTC, Carrilho E, Kist TBL. A review of DNA sequencing techniques. *Q Rev Biophys.* 2002;35(2):169–200.
6. Yang Y, Xie B, Yan J. Application of next-generation sequencing technology in forensic science. *Genomics Proteomics Bioinformatics.* 2014;12:190–197.
7. Mardis ER. Next-generation DNA sequencing methods. *Annu Rev Genomics Hum Genet.* 2008;9:387–402.
8. Morozova O, Marra MA. Applications of next-generation sequencing technologies in functional genomics. *Genomics.* 2008;92(5):255–264.
9. Quail MA, Smith M, Coupland P, et al. A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC Genomics.* 2012;13(341):1–13.
10. Branton D, Deamer DW, Marziali A. The potential and challenges of nanopore sequencing. *Natl Biotechnol.* 2008;26(10):1146–1153.
11. Schneeberger K. Using next-generation sequencing to isolate mutant genes from forward genetic screens. *Nat Rev Genet.* 2014;15:662–676.
12. Lasken RS, McLean JS. Recent advances in genomic DNA sequencing of microbial species from single cells. *Nat Rev Genet.* 2014;15:577–584.
13. Shendure J, Ji H. Next-generation DNA sequencing. *Nat Biotechnol.* 2008;26(10):1135–1145.
14. Linacre A, Templeton JEL. Forensic DNA profiling: state of the art. *Res Rep Forensic Med Sci.* 2014;4:25–36.

Research and Reports in Forensic Medical Science

Publish your work in this journal

Research and Reports in Forensic Medical Science is an international, peer-reviewed, open access journal publishing original research, reports, reviews and commentaries on all areas of forensic medical science. The manuscript management system is completely online and includes a

very quick and fair peer-review system. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.

Submit your manuscript here: <http://www.dovepress.com/research-and-reports-in-forensic-medical-science-journal>

Dovepress