Proceedings of the 5th International Conference on Biological Engineering and Medical Science
DOI: 10.54254/2753-8818/99/2025.22532

From Sanger to Single Molecule: A Comparative Review of DNA Sequencing Technologies and Their Applications

Yi Zhou

Department of Science, Xi'an Jiaotong Liverpool University, Suzhou, China
Yi.Zhou2302@student.xjtlu.edu.cn

Abstract: This review provides a comprehensive overview of DNA sequencing technologies, including Sanger sequencing (the first generation), next-generation sequencing (NGS), and third-generation sequencing (TGS) such as SMRT and Oxford Nanopore MinION. Sanger sequencing, renowned for its high accuracy (up to 99.999% raw per-base accuracy), has limitations in parallel sample analysis, cost-effectiveness, and speed. As exemplified by Illumina's technology, NGS offers high throughput through in situ amplification on a solid surface but has shorter read lengths. TGS technologies like SMRT and Oxford Nanopore MinION have unique advantages: SMRT can sequence previously challenging regions such as AT- or GC-rich, repetitive, and palindromic sequences without bias, while Oxford Nanopore MinION provides ultra-long reads and portability. However, both TGS technologies require high-quality, intact DNA/RNA molecules, posing challenges with cancer samples. Applications of these sequencing technologies are also explored, including direct detection of DNA methylation using SMRT, studying Fragile X gene repeat sequences, crop improvement, and detecting mutations in the GBA gene causing Gaucher disease. Understanding these technologies' applications is crucial for advancing genomics research and related fields.

Keywords: DNA sequencing, Sanger sequencing, Next generation sequencing, Third generation sequencing, SMRT, Oxford Nanopore MinION.

1. Introduction

DNA sequencing is a cornerstone of modern biological research, enabling the determination of the precise order of nucleotides in DNA molecules. Since the development of the first DNA sequencing method, the field has witnessed remarkable progress with the emergence of multiple generations of sequencing technologies. The first generation, represented by Sanger sequencing, has been widely used for decades and has contributed significantly to many biological discoveries. However, its limitations in throughput, cost, and speed have spurred the development of next-generation sequencing (NGS) technologies. NGS platforms, such as Illumina, have revolutionized genomics research by offering high throughput sequencing capabilities, allowing for the analysis of large numbers of samples and genomes in a relatively short time. More recently, third-generation sequencing (TGS) technologies have emerged, providing unique features such as long-read capability and the ability to sequence challenging regions that were difficult to analyze with previous methods. This review aims to provide an overview of Sanger sequencing, NGS, and TGS technologies, including their mechanisms, advantages, and limitations. Additionally, it will discuss the diverse

DOI: 10.54254/2753-8818/99/2025.22532

applications of these sequencing technologies in various fields such as epigenetics, human disease research, and crop improvement, highlighting their potential to drive scientific progress and innovation.

2. Sanger sequencing

2.1. Principle and mechanism

Sanger sequencing, also known as the chain-termination method, is the first-generation DNA sequencing technique developed by Frederick Sanger. This method primarily relies on the ability of dideoxyribonucleotide triphosphates (ddNTPs) to terminate DNA synthesis during replication [1]. In Sanger sequencing, DNA fragments—typically from phage restriction digests—are processed and mixed with primers. The reaction system has DNA polymerase, a mixture of standard deoxynucleotide triphosphates (dNTPs) and fluorescently labeled ddNTPs. The incorporation of a ddNTP terminates strand elongation, resulting in DNA fragments of varying lengths. After labeling and chain extension, the products are denatured and separated by electrophoresis [2]. Unlike traditional slab-gel electrophoresis, modern Sanger sequencing employs capillary electrophoresis, which significantly increases the speed, resolution, and throughput of DNA fragment separation [3].

2.2. Limitations and advantages

A major limitation of Sanger sequencing is the limited scalability due to the small number of samples that can be analyzed in parallel. This limitation arises mainly because Sanger sequencing requires gel electrophoresis, which can separate only a limited number of samples per run [4]. Moreover, Sanger sequencing is not cost-effective for large-scale studies, making it suitable primarily for small subsets of genes sequenced using this method. Additionally, the method is relatively slow, which further restricts its application in scenarios where high-throughput and rapid results are needed [5]. On the other hand**,** the most significant advantage of Sanger sequencing is its high accuracy. After three decades of continuous refinement, the Sanger biochemistry method has been optimized to achieve read lengths of approximately 1,000 bp, with raw per-base accuracy reaching up to 99.999% [6].

3. Next generation sequencing

3.1. Illumina sequencing methodology

Before sequencing, DNA libraries undergo clonal amplification via "bridge amplification" PCR on the sequencing flow cell. Sequencing is performed by detecting fluorescent nucleotides incorporated by DNA polymerase with a reversible terminator. In each cycle, a labeled dNTP is added, its fluorescence is captured, and then the terminator and dye are removed for the next incorporation. This cyclic process is repeated across millions of DNA clusters in parallel. An animation of this process is available on the Illumina website [7].

3.2. Advantages and limitations of NGS

One difference between Sanger sequencing and next-generation sequencing (NGS) is that NGS does not require sequencing reactions to be carried out in individual wells of a microtiter plate. Instead, library fragments are immobilized and amplified in situ on a solid surface, such as a bead or a glass microfluidic channel modified with adapter sequences complementary to the library fragments. This allows Illumina platforms to achieve high throughput, as thousands to millions of library fragments can be sequenced simultaneously. Another key distinction lies in read length: Sanger sequencing is

DOI: 10.54254/2753-8818/99/2025.22532

influenced by gel factors, while NGS is limited by the signal-to-noise ratio, resulting in shorter reads [8].

4. Third generation sequencing

4.1. SMRT sequencing

SMRT, or Single Molecule Real-Time sequencing, was developed by Pacific Biosciences and represents a significant innovation in the field of genomics. Its core mechanism is observing fluorescently labeled nucleotide incorporations. In a designed zero-mode waveguide (ZMW), a single DNA polymerase is fixed at the bottom of a nanowell. As nucleotides with different fluorescent labels are added to the DNA strand by the polymerase during synthesis, they emit distinct-colored flashes of light. A sensitive optical system captures these signals in real time. From the flash colors and sequences, the DNA base sequence is determined, and kinetic parameters, such as the polymerase's reaction rate, can also be extracted [9]. In the research regarding incorporation events, scientists utilized ZMW, a nanophotonic structure that reduces the observation volume by more than three orders of magnitude compared to confocal fluorescence microscopy. Compared to confocal fluorescence microscopy, the ZMW minimizes the observation volume by over three orders of magnitude. This allows for single-molecule analysis at higher concentrations. For instance, in observing DNA polymerase activity, the ZMW's small volume reduces background noise, enabling clear detection of fluorescence bursts from nucleotide incorporation, which benefits biological research and relevant applications [10].

4.2. Oxford Nanopore MinION

The most recent third-generation technology was released by Oxford Nanopore Technologies in 2014 [11]. Their current device, the Oxford Nanopore MinION, is a handheld, portable sequencer that analyzes DNA by electronically measuring the minute disruptions to electric current as DNA molecules pass through a nanopore. The read lengths of the currently available instrument are similar to those produced by PacBio, however, the platform has historically suffered from lower base-calling accuracy and throughput, which has limited its use to sequencing small genomes—such as E. coli (4.5Mbp) or yeast (12Mbp), or amplicons. Using error correction algorithms like those that are available for PacBio reads, the per-nucleotide accuracy of genomes sequenced using the MinION has been measured to be >99.95% [12].

4.3. Comparison between SMRT and MinION

Many sequencing systems struggle with AT- or GC-rich regions, repetitive sequences, or long homopolymers, often producing poor-quality sequences or failing altogether. However, SMRT sequencing overcomes these issues, providing stable sequencing even in previously challenging regions, and can fill the "hard stops" in other systems. An extreme example that can demonstrate the absence of bias in SMRT sequencing is the sequencing of thousands of bases with 100% GC content: the CGG trinucleotide repeat expansion responsible for Fragile X syndrome [13].

Similarly, palindromic sequences—which are symmetrical DNA regions prone to forming secondary structures—are typically lost during the amplification step in most sequencing platforms [6]. Palindromic sequences can't be sequenced in other sequencing systems because, during the amplification stage of sample preparation in those systems, the unique symmetrical structure of such DNA causes it to form secondary structures that interfere with the amplification process and lead to its loss, thus leaving these regions unsequenceable. However, due to the absence of an amplification

DOI: 10.54254/2753-8818/99/2025.22532

step in SMRT sample preparation, palindromic sequences can be readily and accurately sequenced [9].

4.4. General limitations of TGS

Both PacBio HiFi and ONT, two TGS technologies, the two primary third-generation sequencing platforms, offer distinct advantages: PacBio HiFi provides high sequencing accuracy, while ONT generates ultra-long reads and is scalable across different sequencers. However, both require intact, high-molecular-weight DNA or full-length RNA, which can be problematic in cancer research and clinical diagnostics. Many clinical tissue samples—particularly formalin-fixed, paraffin-embedded (FFPE) specimens—are highly fragmented and chemically modified, limiting their compatibility with TGS platforms. One potential solution is the biobanking of fresh frozen tissues which preserves nucleic acid integrity and improves sequencing success [14]. Despite the current limitations of PacBio HiFi and ONT in handling cancer and paraffin-embedded samples, future prospects are promising. Innovate extraction/repair tech, promote fresh frozen tissue biobanking, enhance sequencing platforms, and combine with other genomic techniques. This will overcome limitations, benefiting genomic research and cancer care.

5. Applications of various sequencing types

5.1. Epigenetic studies

This method for detecting DNA methylation relies on the application of zero-mode waveguides (ZMWs) in single-molecule real-time (SMRT) sequencing. ZMWs can determine the incorporation order of fluorescently labeled nucleotides, enable single-fluorophore detection, and provide kinetic information on DNA polymerase activity, including insights into DNA secondary structure [9]. Since DNA methylation influences DNA secondary structure, which in turn affects polymerase kinetics—such as movement speed and pause sites—ZMWs can capture these dynamic changes. A study validated this approach by synthesizing DNA templates with different methylation states, revealing that the effects of DNA methylation on polymerase kinetics vary depending on methylation type and are sequence context-dependent. When combined with circular consensus sequencing (CCS), this method enables the identification of epigenetic modifications at the single-molecule level [15].

5.2. Genetic disease research (fragile X syndrome)

The FMR1 gene contains a CGG repeat in its 5′-untranslated region, which can become unstable when passed to the next generation. In the general population, the repeat length is typically fewer than 55 CGG units, while in individuals with fragile X syndrome (FXS), the repeat length exceeds 200 [16]. Due to the reliance of Sanger sequencing and most "next-generation" sequencing technologies on reading signals from large DNA populations, they tend to lose allelic phase consistency in GC-rich sequences and suffer from reduced resolution as DNA length increases [13,17]. As a result, sequencing of FMR1 alleles with more than approximately 100 CGG repeats is usually not possible, falling short of detecting the full mutation range responsible for FXS. SMRT technology overcomes the limitations by capturing single-molecule signals, allowing it to handle highly repetitive sequences (such as trinucleotide repeat sequences), especially by measuring single molecule signals, thus avoiding issues of sample heterogeneity (phase consistency) and the resolution loss inherent in bulk sequencing. By utilizing Single Molecule Real-Time (SMRT) sequencing to analyze the CGG repeat region of the FMR1 gene, sequencing data for FMR1 alleles with more than 750 CGG repeats can be generated, equivalent to over 2.25 kb of 100% CGG repeat DNA. The repeat size distribution

DOI: 10.54254/2753-8818/99/2025.22532

produced by this method reflects the expected distribution of input DNA (e.g., cloned or PCR-amplified DNA) [17].

5.3. Crop breeding

Genes governing simple, Mendelian traits can be identified by DNA sequencing technologies. Studying their proteins helps us understand traits and potential improvements. These genes can be introduced into cultivars via genetic engineering or through marker-assisted selection (MAS) [18]. Marker-assisted selection (MAS) can accelerate trait selection but is most effective for simple Mendelian traits. Genomic selection (GS), leveraging high-resolution markers and decreasing sequencing costs, enables more efficient breeding, particularly for complex traits. Studies in maize and rice demonstrate the potential of genomic selection in identifying quantitative trait loci (QTLs) for key agronomic traits [19]. In the study of seed color genes in Brassica napus, molecular marker techniques such as SRAP, SCAR, and SNP were used to develop markers tightly linked to the seed color genes. During the process of identifying these markers, sequencing analysis was carried out on the related genes or markers. For example, the tightly linked SRAP marker SA12BG18388 was sequenced using chromosome walking technology, and a 1282 - bp sequence (GenBank accession number EF608928) was obtained. Through the sequencing and flanking sequence analysis of another marker SA7BG29245, 24 SNP loci were discovered. These sequencing efforts provided a basis for the development of stable and efficient molecular markers. By using these markers, the simple Mendelian trait of seed color can be accurately screened in the early stage of breeding. This has accelerated the breeding process of yellow-seeded lines in Brassica napus, fully demonstrating the effectiveness of MAS in the selection of simple traits [20].

5.4. Clinical genetics

The GBA gene encodes glucocerebrosidase. Biallelic mutations in the GBA gene cause Gaucher disease. Heterozygous mutations are risk factors for Parkinson's disease and are also associated with diseases such as dementia with Lewy bodies. The presence of nearby pseudogenes complicates the sequencing of the GBA gene. Therefore, accurately detecting GBA mutations is of great significance for the diagnosis, treatment, and research of related diseases. 102 individuals' GBA genes were selected for sequencing using Oxford Nanopore MinION technology. High-quality genomic DNA was extracted from blood, saliva, or tissue. Sequencing libraries were prepared with specific primers for GBA gene amplification. The amplified products underwent end-repair, adapter ligation, and barcode addition for multiplexing. The MinION device generated abundant sequencing data. Analysis revealed various mutations, including single-nucleotide substitutions, insertions, deletions in coding regions, and intronic SNPs. Advanced algorithms determined the zygosity of missense mutations, yet the RecNciI mutation was missed, perhaps due to complex sequences or algorithm limits. Non-coding SNVs were also detected, and strict quality control identified false positives. NGMLR software reduced false positives in read alignment compared to other tools. Adjusted Nanopolish quality scores effectively distinguished true and false positives, with scores varying by mutation type. A 55 bp exonic deletion was also detected, and haplotype phasing—using advanced algorithms and linkage information, determined allelic relationships, revealing insights into GBA gene haplotypes. Subsequent multiplexing detection showed multiple missense mutations, but one recombinant allele was missed, likely because of its unique sequence during PCR or data analysis, highlighting the challenges of detecting all genetic variations [21].

DOI: 10.54254/2753-8818/99/2025.22532

6. Conclusion

This review summarized the development, principles, advantages, limitations, and applications of three generations of DNA sequencing technologies. Sanger sequencing, while highly accurate, is limited by low throughput and high costs. NGS technologies like Illumina provide high efficiency and throughput but shorter read lengths. TGS platforms such as PacBio SMRT and Oxford Nanopore MinION overcome these limitations by offering long reads and the ability to sequence difficult genomic regions.

These technologies have been applied in areas such as disease research (e.g., Fragile X syndrome, Gaucher disease), epigenetics, and crop improvement. Choosing the right sequencing method is essential for advancing research and practical applications.

Looking ahead, future developments are likely to focus on increasing sequencing accuracy (e.g., improving base-calling algorithms for Nanopore), enhancing the ability to sequence degraded DNA (especially from clinical FFPE samples), and integrating multi-omics data (genomics, epigenomics, transcriptomics) for comprehensive biological insights. Additionally, real-time and point-of-care sequencing using portable devices may expand precision medicine, while AI-driven tools will further optimize data analysis and interpretation. As costs decrease and workflows simplify, whole-genome and methylation sequencing may become standard in both research and clinical diagnostics.

References

[1] *Atkinson, M.R., Deutscher, M.P., Kornberg, A., Russell, A.F. and Moffatt, J.G., 1969. Enzymic synthesis of deoxyribonucleic acid. XXXIV. Termination of chain growth by a 2', 3'-dideoxyribonucleotide. Biochemistry, 8(12), pp.4897-4904.*

[2] *F. Sanger, S. Nicklen, & A.R. Coulson, DNA sequencing with chain-terminating inhibitors, Proc. Natl. Acad. Sci. U.S.A. 74 (12) 5463-5467, https://doi.org/10.1073/pnas.74.12.5463 (1977).*

[3] *Swerdlow H, Gesteland R. 1990. Capillary gel electrophoresis for rapid, high resolution DNA sequencing. Nucleic Acids Res 18:1415–9*

[4] *Ansorge, W.J., 2009. Next-generation DNA sequencing techniques. New biotechnology, 25(4), pp.195-203.*

[5] *Dey, P., 2023. Sanger sequencing and next generation gene sequencing: Basic principles and applications in pathology. In Basic and advanced laboratory techniques in histopathology and cytology (pp. 247-261). Singapore: Springer Nature Singapore.*

[6] *Shendure, J., & Ji, H. (2008). Next-generation DNA sequencing. Nature Biotechnology, 26(10), 1135–1145. doi:10.1038/nbt1486*

[7] *Hu, T., Chitnis, N., Monos, D., & Dinh, A. (2021). Next-generation sequencing technologies: An overview. Human immunology, 82(11), 801-811.*

[8] *Mardis, E. R. (2013). Next-generation sequencing platforms. Annual review of analytical chemistry, 6(1), 287-303.*

[9] *Eid, J., Fehr, A., Gray, J., Luong, K., Lyle, J., Otto, G., ... & Turner, S. (2009). Real-time DNA sequencing from single polymerase molecules. Science, 323(5910), 133-138.*

[10] *Levene, M. J., Korlach, J., Turner, S. W., Foquet, M., Craighead, H. G., & Webb, W. W. (2003). Zero-mode waveguides for single-molecule analysis at high concentrations. science, 299(5607), 682-686.*

[11] *Lee, H., Gurtowski, J., Yoo, S., Nattestad, M., Marcus, S., Goodwin, S., Richard McCombie, W. and Schatz, M.C., 2016. Third-generation sequencing and the future of genomics. BioRxiv, p.048603.*

[12] *Loman, N.J., Quick, J. & Simpson, J.T. A complete bacterial genome assembled de novo using only nanopore sequencing data. Nat Methods (2015)*

[13] *Korlach, J., & Biosciences, P. (2013). Understanding accuracy in SMRT sequencing. Pac Biosci, 2013, 1-9.*

[14] *Ermini, L. and Driguez, P., 2024. The application of long-read sequencing to cancer. Cancers, 16(7), p.1275.*

[15] *Flusberg, B. A., Webster, D. R., Lee, J. H., Travers, K. J., Olivares, E. C., Clark, T. A., ... & Turner, S. W. (2010). Direct detection of DNA methylation during single-molecule, real-time sequencing. Nature Methods, 7(6), 461-465.*

[16] *Willemsen, R., Levenga, J., & Oostra, B. A. (2011). CGG repeat in the FMR1 gene: size matters. Clinical genetics, 80(3), 214-225.*

[17] *Loomis, E. W., Eid, J. S., Peluso, P., Yin, J., Hickey, L., Rank, D., ... & Hagerman, P. J. (2013). Sequencing the unsequenceable: expanded CGG-repeat alleles of the fragile X gene. Genome Research, 23(1), 121-128.*

[18] *Edwards, D. and Batley, J. (2010), Plant genome sequencing: applications for crop improvement. Plant Biotechnology Journal, 8: 2-9. https://doi.org/10.1111/j.1467-7652.2009.00459.x*

[19] *Jackson, S.A., Iwata, A., Lee, S.H., Schmutz, J. and Shoemaker, R., 2011. Sequencing crop genomes: approaches and applications. New Phytologist, 191(4), pp.915-925.*

[20] *Rahman, M. M. (2007). Development of molecular markers for marker assisted selection for seed quality traits in oilseed rape.*

[21] *Leija‐Salazar, M., Sedlazeck, F.J., Toffoli, M., Mullin, S., Mokretar, K., Athanasopoulou, M., Donald, A., Sharma, R., Hughes, D., Schapira, A.H. and Proukakis, C., 2019. Evaluation of the detection of GBA missense mutations and other variants using the Oxford Nanopore MinION. Molecular genetics & genomic medicine, 7(3), p. e564.*