



From First-Generation to Third-Generation DNA Sequencing: Technologies and Applications in Agriculture

*Jay M. Khaniya, Ashish G. Vala, Manoj V. Parakhia, Mahesh B. Vaja, Darshana G. Hirpara and Shraddha B. Bhatt

College of Agriculture, Junagadh Agricultural University, Junagadh – 362001

*Corresponding Author's email: jaykhaniya@gmail.com

DNA sequencing has transformed biological research by enabling scientists to read the genetic information encoded in deoxyribonucleic acid (DNA). From the chain-termination method developed by Frederick Sanger and colleagues to modern high-throughput short-read and long-read sequencing platforms, technological progress has dramatically increased sequencing capacity, improved genome reconstruction and expanded access to genomic information. The evolution of DNA sequencing is commonly described in three generations: first-generation sequencing, represented by Sanger sequencing; second-generation sequencing, characterised by massively parallel sequencing of millions of DNA fragments; and third-generation sequencing, which enables the analysis of individual DNA molecules and produces long reads. Major platforms include the ABI PRISM and Applied Biosystems capillary sequencers, Roche 454, Illumina MiSeq and NovaSeq, Ion Torrent, Pacific Biosciences (PacBio) SMRT systems, and Oxford Nanopore Technologies' MinION and PromethION. These technologies differ in read length, throughput, accuracy, cost, sample requirements, and analytical capabilities. Their applications in agriculture range from marker discovery and genetic diversity analysis to genome assembly, disease-resistant gene identification, genomic selection, crop pangenomics, and precision breeding. Understanding the strengths and limitations of each generation is essential for selecting appropriate sequencing strategies for agricultural research. This article reviews the evolution of DNA sequencing technologies, representative sequencer models, and their implications for sustainable crop improvement.

Introduction: Reading the Language of Life

Every crop plant carries a genetic blueprint that influences its growth, development, productivity, and response to environmental stresses. This blueprint is encoded in DNA, a molecule composed of four nitrogenous bases: adenine (A), thymine (T), guanine (G), and cytosine (C). The order of these bases stores biological information that determines gene structure, gene regulation, and inherited variation.

For decades, scientists have attempted to understand how differences in DNA sequences contribute to variation in important agricultural traits, including grain yield, seed quality, drought tolerance, salinity tolerance, and disease resistance. The principal challenge was not merely to extract DNA but to read its nucleotide sequence accurately and efficiently.

The development of DNA sequencing fundamentally changed this situation. What initially required laborious analysis of a limited number of DNA fragments has evolved into automated systems capable of reading millions to billions of bases in a sequencing run. Modern platforms can also generate long DNA reads, helping researchers resolve repetitive genomic regions, structural variations, and complex genome arrangements.

The importance of this evolution extends far beyond the laboratories of molecular biology. In agricultural research, sequencing provides opportunities to discover genes

associated with desirable traits, identify molecular markers, characterise crop diversity, and support marker-assisted selection and genomic selection.

The evolution of sequencing is commonly organised into three broad generations. However, this classification is a useful framework rather than a rigid boundary: different technologies have overlapping capabilities, and newer systems continue to improve the accuracy, throughput, and accessibility of existing approaches.

First Generation: The Foundation of DNA Sequencing

The development of Sanger sequencing

The first major breakthrough in DNA sequencing came in 1977, when Frederick Sanger, Steven Nicklen, and Alan Coulson described a DNA sequencing method based on chain-terminating inhibitors. Their approach, commonly called the dideoxy chain-termination method, provided a practical way to determine the nucleotide sequence of DNA.

The method uses DNA polymerase to synthesise new DNA strands in the presence of normal deoxynucleotides and modified nucleotides called dideoxynucleotides (ddNTPs). When a ddNTP is incorporated into a growing DNA strand, further extension stops. This produces DNA fragments of different lengths, each ending at a particular nucleotide.

In the classical method, the fragments were separated by electrophoresis and the sequence was reconstructed from their order. Later developments introduced fluorescent labelling, automated fragment detection, and capillary electrophoresis, substantially improving the speed and practicality of the technique.

The original chain-termination method was a landmark in molecular genetics and contributed to the development of automated DNA sequencing and large-scale genome projects.

Representative sequencers

Important instruments associated with the development of automated Sanger sequencing include:

- **Applied Biosystems ABI PRISM 310 Genetic Analyser:** An early automated capillary electrophoresis instrument that helped bring fluorescent DNA analysis into routine laboratories.
- **ABI PRISM 3700 DNA Analyser:** A high-throughput capillary sequencing system used in large-scale sequencing projects, including genome research.
- **ABI PRISM 3730 and 3730xl DNA Analysers:** Widely used capillary electrophoresis systems for DNA sequencing and fragment analysis.

These instruments represent the automation of first-generation sequencing. Their importance lies in the integration of fluorescent detection, capillary separation, and computerised sequence interpretation.

Advantages and limitations

Sanger sequencing is valued for its high accuracy in suitable applications, relatively straightforward interpretation, and ability to produce reads commonly around 500 - 1,000 bases, depending on the instrument, chemistry, and template quality.

Its principal limitation is throughput. Each sequencing reaction generally examines a particular DNA fragment, making the method inefficient for sequencing entire large genomes or thousands of samples compared to massively parallel technologies.

Agricultural applications

Despite the arrival of newer sequencing methods, Sanger sequencing remains useful in agricultural biotechnology for the following reasons:

- Confirmation of PCR products and cloned DNA fragments.
- Verification of the identity of molecular markers and candidate genes.
- Validation of selected single-nucleotide polymorphisms (SNPs) and mutations.
- Confirmation of plasmid constructs and gene-editing outcomes.
- Checking the sequence of newly designed primers or targeted DNA regions.

The key contribution of first-generation sequencing was the establishment of a reliable DNA reading as a practical laboratory technique. Its legacy remains visible in routine molecular biology, plant genetics, and sequence validation.

Second Generation: Massively Parallel DNA Sequencing

The next-generation sequencing revolution

Second-generation sequencing, commonly called next-generation sequencing (NGS), transformed DNA analysis by allowing millions of DNA fragments to be sequenced simultaneously.

Instead of processing a small number of fragments independently, these technologies prepare a DNA library, distribute the fragments across a sequencing surface or reaction system, and analyse large numbers of sequencing events in parallel.

Most second-generation platforms use amplified DNA templates, although the exact chemistry differs between systems. Their defining contribution is massively parallel sequencing, which substantially increases the volume of sequence data generated per run.

Several important technologies emerged during this period, including pyrosequencing, sequencing by ligation, semiconductor sequencing, and sequencing by synthesis.

Major second-generation platforms

A. Roche 454 GS FLX: Pyrosequencing

The Roche 454 platform was an early commercial next-generation sequencing system. The method used pyrosequencing, in which nucleotide incorporation was detected through a biochemical reaction that generates light.

Its relatively long reads compared with many early short-read systems made it useful for applications such as microbial genome sequencing, amplicon analysis, and early genome assembly.

However, the technology had difficulty accurately determining the length of long stretches of identical nucleotides, known as homopolymers. Roche 454 sequencing was subsequently discontinued, but played an important historical role in establishing massively parallel sequencing.

B. Illumina sequencing: Sequencing by synthesis

Illumina developed one of the most widely adopted approaches to second-generation sequencing. Its sequencing-by-synthesis technology detects fluorescent signals as labelled nucleotides are incorporated into growing DNA strands.

In a typical workflow, DNA is fragmented and prepared in a sequencing library. Library fragments bind to a flow cell, undergo cluster generation, and are sequenced through successive cycles of nucleotide incorporation and imaging. The resulting signals are converted into nucleotide sequences by computational base calling.

Representative Illumina platforms include:

- **MiSeq:** A benchtop instrument suitable for targeted sequencing, small genomes, amplicons, and selected microbiome studies.
- **NextSeq series:** medium to high-throughput systems suitable for a range of genomic and transcriptomic applications.
- **NovaSeq series, including NovaSeq X:** High-throughput platforms designed for large-scale sequencing projects and population genomics.

The capabilities of individual models vary, and the names above represent platform families rather than identical specifications. Illumina sequencing is particularly useful for SNP discovery, genotyping-by-sequencing, RNA sequencing, genome resequencing, genome-wide association studies, and genomic selection. Its major strength is the generation of large volumes of accurate short-read data. A limitation is that short reads can be difficult to assemble or map correctly in highly repetitive genomic regions.

C. Applied Biosystems SOLiD: Sequencing by ligation

The SOLiD system used sequencing-by-ligation chemistry. Instead of relying primarily on polymerase-mediated nucleotide incorporation, it used DNA ligase and labelled oligonucleotide probes to infer sequence information. The technology was known for its

distinctive colour-space encoding and high accuracy in certain applications. However, its workflow and data interpretation differed from those of conventional sequencing-by-synthesis platforms, and SOLiD is now primarily of historical importance.

D. Ion Torrent: Semiconductor sequencing

Ion Torrent introduced a different approach by detecting hydrogen ions released during nucleotide incorporation. These changes in hydrogen-ion concentration produce a measurable signal on a semiconductor chip.

Representative instruments include the Ion PGM and Ion GeneStudio S5 systems. Ion Torrent platforms can support targeted sequencing, amplicon sequencing, and selected microbial or genomic studies. They do not require fluorescence imaging to detect nucleotide incorporation.

A major technical consideration is that estimating the number of nucleotides incorporated during a single event can be challenging in long homopolymer stretches.

Advantages and limitations of second-generation sequencing

The major advantages of second-generation sequencing include the following:

1. **High throughput:** Large numbers of DNA fragments can be sequenced simultaneously.
2. **Cost efficiency on scale:** The cost per base can be low when sufficient sequencing capacity is used.
3. **Accurate short-read data:** These data are highly useful for SNP detection and many forms of genotyping.
4. **Broad applicability:** The technology supports genome resequencing, transcriptomics, metagenomics, and targeted analysis.
5. **Population-level studies:** Large collections of crop accessions can be analysed for genetic variation.

The main limitations include difficulty resolving long repetitive sequences, complex structural variations, and some regions of highly similar sequence. The short fragments must often be assembled computationally, and the results depend on library quality, sequencing depth, reference-genome quality, and bioinformatics methods.

Agricultural applications

Second-generation sequencing has become an important resource for crop improvement. In plant breeding, it enables researchers to discover thousands or millions of SNPs, develop molecular markers, construct linkage maps, and identify genomic regions associated with quantitative traits. Genotyping-by-sequencing and related reduced-representation approaches can help characterise germplasm and support genome-wide association studies.

For example, a researcher studying drought tolerance in groundnuts could use sequencing-based genotyping to identify SNPs associated with performance under water-limited conditions. Candidate markers could then be evaluated in independent breeding populations before being incorporated into selection programmes.

Similarly, in pigeonpea, sequencing can support genetic diversity analysis, association mapping, and the investigation of genomic regions linked to resistance against important diseases.

The crucial point is that sequencing identifies genetic information; demonstrating that a marker is consistently associated with a useful trait requires suitable populations, sound experimental design, and validation across relevant environments.

Third Generation: Single-Molecule and Long-Read Sequencing

Why do long reads matter

Although second-generation sequencing produces large quantities of sequence data, short reads can be inadequate to reconstruct complex genomes. Many plant genomes contain repetitive DNA, duplicated genes, transposable elements, and structural variations that are difficult to resolve using short fragments alone.

Third-generation sequencing approaches address this problem by reading individual DNA molecules, generally without requiring clonal amplification of each template as part of the sequencing reaction.

Two major technologies dominate this field: single-molecule real-time sequencing developed by Pacific Biosciences (PacBio) and nanopore sequencing developed commercially by Oxford Nanopore Technologies (ONT). These approaches are not identical. PacBio detects DNA polymerase activity in real time, whereas nanopore sequencing detects changes in electrical current as nucleic acids pass through a nanopore.

Pacific Biosciences: SMRT sequencing

Pacific Biosciences introduced single-molecule real-time (SMRT) sequencing, which observes DNA synthesis by an individual DNA polymerase in a specialised observation chamber. During sequencing, fluorescent signals associated with nucleotide incorporation are detected and interpreted to determine the DNA sequence. A major development within this technology is **HiFi sequencing**. In this approach, circular DNA templates are read repeatedly, allowing a consensus sequence to be generated from multiple observations of the same molecule. This combines long reads with very high sequence accuracy.

Representative PacBio instruments include:

- **PacBio RS II:** An earlier SMRT sequencing system that helped establish long-read sequencing.
- **Sequel and Sequel II/IIe:** Systems that expanded throughput and supported a broad range of long-read applications.
- **Revio:** A high-throughput platform designed for large-scale HiFi sequencing.
- **Vega:** A more compact PacBio system designed to make HiFi sequencing accessible to a wider range of laboratories.

Instrument availability and specifications change over time; therefore, historical and current models should not be treated interchangeably.

PacBio sequencing is especially useful for de novo genome assembly, structural-variant detection, haplotype resolution, transcript analysis, and the reconstruction of complex plant genomes. Depending on the assay, it can also provide information about the DNA base modifications.

Its limitations include the need for suitable high-molecular-weight DNA, specialised library preparation, instrument and consumable costs, and computational resources for analysing large datasets.

Oxford Nanopore Technologies: Nanopore sequencing

Oxford Nanopore Technologies developed a sequencing approach in which DNA or RNA passes through a nanoscale pore embedded in a membrane. As the nucleic acid moves through the pores, its sequence-dependent physical properties produce characteristic changes in the electrical current. These signals are interpreted by computational base-calling systems to infer the nucleotide sequence.

One of the distinctive features of nanopore sequencing is that the read length is strongly influenced by the length and quality of the input nucleic acid molecules. Very long reads can be generated when intact, high-molecular-weight DNA is available.

Representative Oxford Nanopore platforms include:

- **MinION:** A portable, compact sequencing device suited for field-orientated research and smaller sequencing projects.
- **GridION:** A benchtop system that can operate multiple MinION flow cells.
- **PromethION:** A scalable platform family designed for higher-throughput sequencing.
- **Flongle:** A smaller flow-cell option designed for lower-throughput experiments, compatible with selected MinION workflows.

Oxford Nanopore technology supports real-time data acquisition, enabling researchers to begin to analyse sequence data while a run is in progress. It can also detect certain DNA modifications directly from native DNA signals when supported by the appropriate chemistry and analysis methods. Its applications include pathogen surveillance, genome assembly, structural-variant detection, metagenomics, transcript analysis, and sequencing in settings where portability is valuable. Important limitations include the dependence of results on flow-cell performance, sample preparation, base-calling models, and analysis pipelines. Accuracy varies with

chemistry, software, and sequencing conditions, so comparisons must use current, application-specific performance data.

PacBio versus Oxford Nanopore

Both technologies generate long reads, but differ in their underlying chemistry and data characteristics.

PacBio HiFi sequencing is particularly attractive when researchers need long reads with very high consensus accuracy. Oxford Nanopore sequencing is notable for its flexible read lengths, real-time analysis, and portable instrument options.

Neither platform is universally superior. The appropriate choice depends on whether the main objective is highly accurate genome assembly, detection of structural variation, field-based pathogen identification, methylation analysis, transcript characterisation, or high-throughput sequencing.

Applications in crop genomics

Long-read sequencing is particularly important for plant species with large, repetitive, or polyploid genomes.

For example, a short-read genome assembly may leave gaps in repetitive regions or fail to resolve duplicate genomic segments. For example, a short-read genome assembly may leave gaps in repetitive regions or fail to resolve duplicate genomic segments. For example, a short-read genome assembly may leave gaps in repetitive regions or fail to resolve duplicate genomic segments. Long reads can bridge these regions, improve chromosome-scale assemblies, and reveal structural variations that might otherwise remain undetected.

In cotton, wheat, and other crops with complex genomes, long-read sequencing can help researchers distinguish closely related genomic regions and investigate variation among cultivars or wild relatives.

In pigeonpea and groundnut research, it can support the reconstruction of reference genomes, identification of candidate resistance genes, and investigation of genomic differences among accessions. These findings can provide a foundation for future markers and breeding studies.

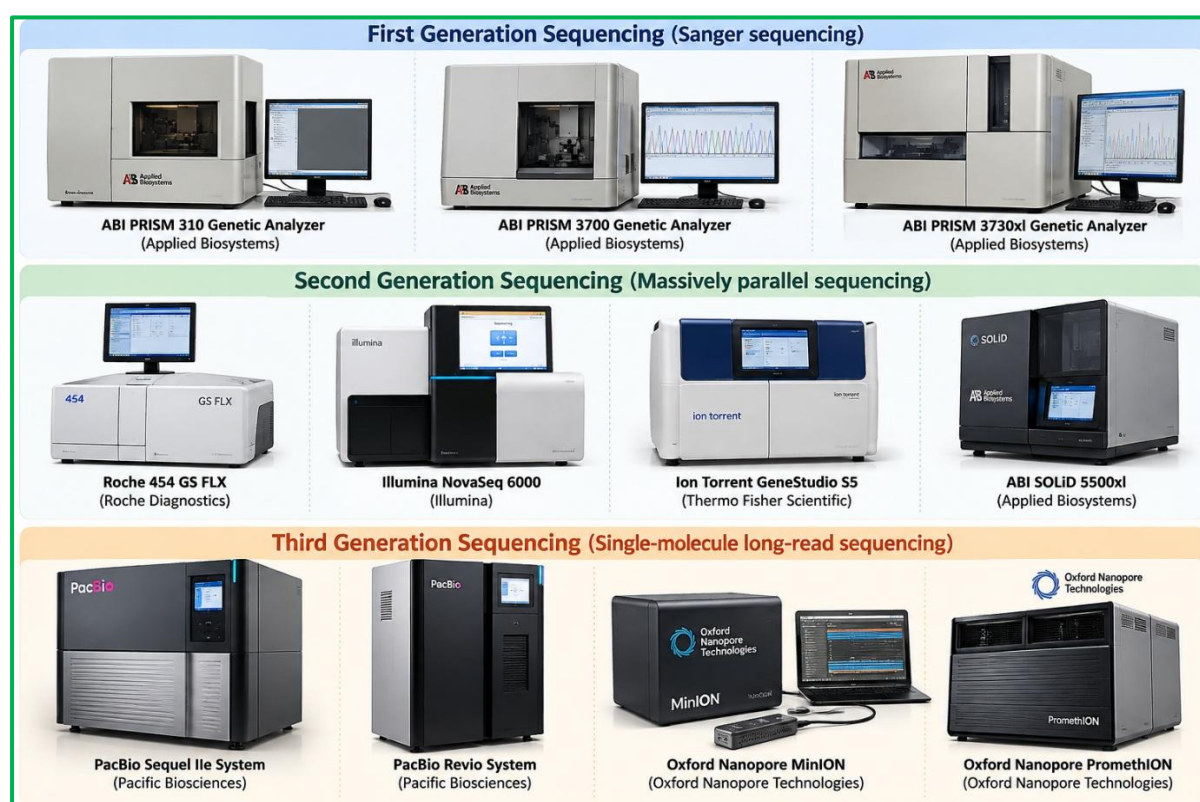
A Comparative Overview of the Three Generations

The following table summarises the main distinctions among the major generations of sequencing technologies.

Feature	First generation	Second generation	Third generation
Representative technology	Sanger chain termination	Massively parallel sequencing	Single-molecule long-read sequencing
Major examples	ABI PRISM 310, 3700, 3730xl	Roche 454, Illumina MiSeq/NovaSeq, Ion Torrent, SOLiD	PacBio RS II, Sequel IIe, Revio, Vega; ONT MinION, GridION, and PromethION
Template amplification	PCR commonly used in preparation; individual reactions analysed	Cluster or template amplification is common, depending on the platform.	Individual-molecule sequencing without clonal amplification in the sequencing reaction
Typical read characteristics	Hundreds of bases, often approaching 1 kb	Generally short reads; exact lengths depend on platform and chemistry	Long reads, potentially tens of kilobases or more

Feature	First generation	Second generation	Third generation
Throughput	Low to moderate	High to very high	Moderate to very high, depending on the platform
Principal strength	Targeted sequence verification and high-quality reads	Large-scale, accurate short-read sequencing	Long-range genomic information and structural-variant resolution
Major limitation	Limited throughput for large projects	Short reads complicate some repetitive and complex regions	Cost, DNA-quality requirements, platform-specific errors, and computational demands
Agricultural applications	Validation of markers and DNA constructs	SNP discovery, genotyping, GWAS and genomic selection	Reference genome assembly, haplotypes, Structural Variations, and pangenomics

Note: The table gives general characteristics rather than universal performance specifications. Read lengths, accuracy, throughput, and cost depend on the instrument, chemistry, library preparation, and analysis method.



Beyond the Three Generations: What Is Next?

The evolution of sequencing is continuing, but it is important not to assume that every emerging technology constitutes an established fourth generation. Several developments are shaping the next phase of genomics:

Higher accuracy and throughput: Improved chemistry, instrument design, and computational methods continue to increase the quality and volume of sequence data.

Improved genome completeness: Long-read sequencing, combined with complementary data and improved assembly algorithms, is making it possible to reconstruct increasingly complete genomes, including regions that were previously difficult to resolve.

Direct detection of DNA modifications: Selected sequencing technologies can identify certain base modifications using signals generated during sequencing, reducing the need for separate assays in some applications.

Pangenomics: A single reference genome cannot represent all the genetic diversity of a crop species. Plant pangenomes combine information from multiple genomes, helping identify genes and structural variants absent from a single reference.

Integration of multiple data types: Genomic, transcriptomic, epigenomic, and phenotypic information can be combined to improve understanding of how genetic variation influences agronomic traits.

These developments are more accurately described as advances in sequencing and genome analysis than as a universally accepted fourth generation of DNA sequencing.

Implications for Agricultural Research and Crop Improvement

The real value of DNA sequencing lies not in the number of bases generated, but in how sequence information is translated into biological understanding and practical outcomes.

Marker-assisted selection

Sequencing can identify polymorphisms associated with genes or genomic regions that control important traits. Once suitable markers are validated, they can help breeders select plants that carry desirable alleles without relying exclusively on phenotypic observations.

For example, validated markers associated with disease resistance can support early-generation screening in breeding populations. Their usefulness depends on the strength of the marker–trait association and its reliability across genetic backgrounds and environments.

Genomic selection

Genomic selection uses genome-wide marker information to estimate the genetic merit of breeding individuals. High-throughput sequencing can contribute to the discovery and generation of marker datasets used to develop genomic prediction models.

When combined with accurate phenotypic records and well-designed training populations, these approaches can help breeders select promising lines earlier in the breeding cycle.

Resistance to diseases and pathogen identification

Sequencing can help characterise plant pathogens, investigate genetic variation in pathogen populations, and identify candidate resistance genes in crops. Long-read sequencing may also improve the reconstruction of complex pathogen genomes.

However, sequencing alone does not establish that a detected pathogen causes a particular disease. Biological confirmation, appropriate controls, and epidemiological evidence remain important.

Climate-resilient agriculture

Drought, salinity, heat, and emerging diseases threaten agricultural productivity. Sequencing can support the identification of genetic variants associated with stress responses and help researchers investigate the mechanisms underlying adaptation.

Combining these genomic findings with field-based phenotyping, multi-environment trials, and breeding programmes can support the development of crops suited to changing climatic conditions.

Conservation of agricultural biodiversity

Traditional cultivars, landraces, and crop wild relatives contain genetic diversity that may be useful for future crop improvement. Sequencing helps characterise this diversity and identify potentially valuable alleles.

Pangenome approaches extend this opportunity by capturing genetic information from multiple accessions rather than relying on a single reference genome.

Choosing the Right Sequencing Technology

The best sequencing platform is the one that addresses the research question with an appropriate balance of accuracy, read length, throughput, cost, and analytical requirements.

For routine validation of a PCR product or candidate gene, Sanger sequencing may be sufficient. For SNP discovery across hundreds of breeding lines, second-generation sequencing may offer a practical and economical approach. For assembling a complex plant genome or resolving structural variations, long-read sequencing may be more suitable.

A combined strategy can often be more effective than relying on a single platform. For example, long reads can provide the continuity required for genome assembly, while accurate short reads can contribute complementary evidence for small-variant analysis or sequence polishing, where appropriate. PacBio HiFi data can also provide long reads with high consensus accuracy, reducing the need for such hybrid approaches in some projects.

Before selecting a platform, researchers should consider the following:

1. The biological question and the intended outcome.
2. Genome size, ploidy, repetitive content, and DNA quality.
3. The number of samples and the required depth of sequencing.
4. The type of variation to be detected.
5. Instrument access, consumable costs, and data-storage requirements.
6. Expertise available in bioinformatics and downstream validation.

Sequencing data should be evaluated through appropriate quality-control procedures, validated analytical pipelines, and independent biological evidence wherever necessary.

Conclusion: From Reading DNA to Improving Crops

The history of DNA sequencing represents a progression from targeted reading of DNA fragments to high-throughput genome analysis and long-range reconstruction of genetic information. First-generation sequencing established the foundations of accurate DNA reading; second-generation sequencing made large-scale genomic analysis practical; and third-generation technologies expanded the ability to study long DNA molecules, complex genome structures, and certain epigenetic features.

For agricultural biotechnology, these advances have opened new opportunities in molecular marker development, disease-resistance research, genetic diversity analysis, genome-assisted breeding, and crop pangenomics.

Nevertheless, no sequencing technology can independently guarantee superior crop variety. Meaningful crop improvement requires the integration of reliable sequence data with sound genetics, rigorous phenotyping, statistical analysis, biological validation, and field-based selection.

The future of agricultural genomics will therefore depend not simply on sequencing more DNA but on using the right technology to answer the right biological question. When sequencing is integrated effectively with plant breeding and agricultural research, it can help transform genetic information into more productive, resilient, and sustainable farming systems.

References

1. Sanger, F., Nicklen, S., & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences*, 74(12), 5463–5467. <https://doi.org/10.1073/pnas.74.12.5463>
2. Bentley, D. R., Balasubramanian, S., Swerdlow, H. P., et al. (2008). Accurate whole human genome sequencing using reversible terminator chemistry. *Nature*, 456, 53–59. <https://doi.org/10.1038/nature07517>
3. Goodwin, S., McPherson, J. D., & McCombie, W. R. (2016). Coming of age: Ten years of next-generation sequencing technologies. *Nature Reviews Genetics*, 17, 333–351. <https://doi.org/10.1038/nrg.2016.49>
4. Heather, J. M., & Chain, B. (2016). The sequence of sequencers: The history of sequencing DNA. *Genomics*, 107(1), 1–8. <https://doi.org/10.1016/j.ygeno.2015.11.003>
5. Van Dijk, E. L., Jaszczyszyn, Y., Naquin, D., & Thermes, C. (2018). The third revolution in sequencing technology. *Trends in Genetics*, 34(9), 666–681. <https://doi.org/10.1016/j.tig.2018.05.008>

6. Schreiber, M., Jayakodi, M., Stein, N., *et al.* (2024). Plant pangenomes for crop improvement, biodiversity and evolution. *Nature Reviews Genetics*, 25, 563–577. <https://doi.org/10.1038/s41576-024-00691-4>
7. Illumina. Sequencing by synthesis technology. <https://www.illumina.com/science/technology/next-generation-sequencing/sequencing-technology.html>
8. Pacific Biosciences. HiFi sequencing technology and sequencing systems. <https://www.pacb.com/technology/hifi-sequencing/how-it-works/>
9. Oxford Nanopore Technologies. DNA and RNA sequencing devices. <https://nanoporetech.com/products/sequence>