Seed Res. 53 (2): 97-104, December, 2025
doi: https://doi.org/10.56093/SR.v53i2.1

Review Paper

Application of Next-Generation Sequencing (NGS) Technologies for Efficient Breeding, Quality Assurance and Testing of Seeds

BONY DE KUMAR*

Yale Center for Genome Analysis, Yale University, New Haven, CT, USA, 06516

*bony.dekumar@yale.edu

(Received August 2025; Revised October 2025; Accepted October 2025)

ABSTRACT: Seed constitutes the foundational resource for global agriculture and food security. The availability of high-quality, genetically pure, and pathogen-free seed is indispensable for ensuring sustainable crop productivity and resilient farming systems. Traditional seed testing approaches—such as morphological characterization, protein or isozyme profiling, and PCR-based molecular assays—continue to provide valuable information but often lack the resolution, scalability, and diagnostic breadth required to address modern seed industry challenges.Next-generation sequencing (NGS) has transformed our ability to interrogate seed biology by enabling high-resolution analysis of seed genomes, transcriptomes, epigenomes, and microbiomes within a single technological framework. This review summarizes the evolution of sequencing platforms, clarifies key terminology, and delineates how advanced sequencing technologies are increasingly integrated into seed research, seed technology innovation, and quality assurance programs. The transition from first-generation Sanger sequencing to second-generation high-throughput sequencing dramatically reduced sequencing cost while exponentially increasing data output. Illumina sequencing, as the dominant short-read system, provided reliable and cost-efficient genome-wide genotyping solutions. The advent of third-generation long-read platforms, including PacBio and Oxford Nanopore technologies, has enabled accurate resolution of repetitive regions, structural variants, and complex polyploid genomes—characteristics that are highly relevant to many modern cultivated crops. Integrating complementary sequencing approaches with massively parallel data acquisition has deepened our understanding of plant genetic architecture and the molecular determinants of yield performance, stress resilience (biotic and abiotic), seed vigor, and post-harvest quality. Emerging technologies—such as advanced nanopore chemistries, spatially resolved and in situ sequencing, and microscopy-coupled nucleic acid profiling—promise to further accelerate discovery and operational deployment. Collectively, these innovations are reshaping the seed sector, providing unprecedented opportunities for precision breeding, rigorous genetic purity testing, comprehensive phytosanitary certification, and the development of next-generation seed quality standards.

Keywords: Next-generation sequencing, Seed quality assurance, Genomic technologies

History of sequencing technologies

Over time, sequencing platforms and technologies have developed in four distinct generations[1-5]. These are defined by their throughput, read length, and sample input type. First-generation sequencing had limitations in the amount and type of DNA sequence it could generate in one run. While highly accurate, it is limited by low throughput and relatively short read lengths (typically 500–900 bp per reaction). Its applications were therefore restricted to sequencing specific gene regions. These regions were amplified and sequenced individually in a low-throughput manner. Sanger sequencing is an example of first-generation sequencing and remains widely used in both research and clinical applications [6]. Despite its limitations, Sanger sequencing remains a gold standard in many research and clinical settings due to its precision and reliability.

The advent of second-generation sequencing in the mid-2000s marked a turning point in genomics. Platforms such as Illumina and Ion Torrent introduced massively parallel sequencing, which simultaneously processes millions of short DNA fragments. These “short reads” (typically 50–300 bp) are generated from individual DNA molecules and then computationally aligned to a reference genome or assembled de novo. Multiple samples can be mixed and sequenced together using a molecular barcoding strategy where a unique 6-12 bp sequence serves as an index for each sample. Reads from samples are separated based on these indices using a computational

approach called demultiplexing. Second-generation technologies enabled whole-genome sequencing (WGS), whole-exome sequencing (WES), transcriptome profiling, and targeted resequencing at dramatically reduced costs compared to Sanger sequencing. Second-generation technologies offer high-throughput platforms and sequencing at a very low cost; however, they cannot be used for the analysis of repeat regions or the identification of structural variants [7-9].

Third-generation sequencing technologies, including PacBio's Single-Molecule Real-Time (SMRT) sequencing and Oxford Nanopore Technologies (ONT), addressed many of the limitations of short-read sequencing. Third-generation sequencing addressed the challenges of assembling complex short-read data and resolving repeat sequences or large genomic rearrangements. This was made possible by long reads (1 kb to 2 Mb) and single-molecule sequencing. Long-read sequencing provides higher certainty in read overlap and assembly, leading to improved resolution of repetitive regions and structural variations, and offers a clearer view of structural variations, repetitive elements, and haplotypes. These platforms provide several advantages over others, including the label-free sequence determination of native DNA and RNA molecules without the need for amplification. These capabilities have been crucial for applications in de novo genome assembly, metagenomics, and plant breeding for species with polyploid genomes [10-14].

The most recent wave of innovation—sometimes referred to as fourth-generation sequencing—has been driven by advances in microfluidics, ultra-low input library preparation, and integrated multi-omics. These methods enable the analysis of genetic, transcriptomic, and epigenetic information from single cells, opening new avenues for understanding cellular heterogeneity and dynamic biological processes.

Glossary of Common Terminology in Next-Generation Sequencing (NGS)

Term	Definition	Notes/Applications
Read	A DNA or RNA sequence generated from a single molecule during sequencing.	Can be short (50–600 bp, Illumina) or long (1 kb–2 Mb, PacBio/ONT).
Read Length	The number of nucleotides in a sequencing read.	
Throughput	Total amount of sequencing data generated in a run (measured in Gb or million clusters).	
Coverage (Depth)	Average number of times a genomic position is sequenced.	Each application needs a define range of coverage. eg WGS need minimum 30X coverage. Short read and long read requires different coverage for same applications.
Library Preparation	Conversion of DNA/RNA into molecule as per the requirement of sequencer.	Involves fragmentation, adapter ligation, and often amplification.
Adapters	Synthetic DNA sequences ligated to fragments for sequencing.	Allow fragments to attach to the flow cell or pores.
Indices (Barcodes)	Unique short sequences added to adaptors to identify individual samples	Enable sequencing of multiple samples in single run in a sequencing lane.
Demultiplexing	Computational process of separating reads back to their original samples using Barcode/indices.	Essential for pooled sequencing experiments.
Paired-End Sequencing	Sequencing from both ends of a DNA fragment.	Provides two reads per fragment; improves alignment and read accuracy.
Single-Molecule Sequencing	Sequencing of native DNA/RNA molecules without amplification.	Reduces PCR bias; enables direct epigenetic modification detection.
De Novo Assembly	Reconstruction of a genome without a reference sequence.	Useful for generating genome assembly for new genomes
Reference Alignment	Mapping reads to a known reference genome.	Used for variant calling, expression quantification, and comparative genomics.
Variant Calling	Identification of sequence variants (SNVs, indels, SVs) compared to a reference genome.	
Structural Variants (SVs)	Large genomic alterations (insertions, deletions, inversions, translocations).	More accurately detected with long-read sequencing.
Base Quality Score (Q-Score)	Confidence score for each base call.	Q30 = 99.9% accuracy; commonly used benchmark for sequencing quality.

Workflow of next-generation sequencing

Next-generation sequencing (NGS) offers comprehensive and high-resolution insights into genomes, transcriptomes, epigenomes, and microbiomes, and its application has expanded rapidly into seed research, technology development, and quality control systems [15, 16]. Central to NGS workflows are library preparation and sequencing, which together determine data quality, reproducibility, and the breadth of biological insight gained. Each methodological choice, ranging from nucleic acid fragmentation to adapter design and sequencing chemistry, has profound implications for downstream analyses. While sequencing platforms continue to evolve, the optimization and standardization of library preparation remain essential for ensuring accuracy and maximizing translational impact.

Library preparation refers to the conversion of DNA or RNA into a format compatible with sequencing. In most cases, RNA must first be converted into complementary DNA (cDNA) before library construction begins. The process typically includes fragmentation, end repair, adapter ligation, amplification, and, when required, enrichment or selection of specific molecules. Each of these steps may introduce biases that affect the representation of the original sample, making library preparation one of the most critical variables in NGS workflows [17, 18].

Fragmentation is required to generate nucleic acid fragments of appropriate size for sequencing. This can be achieved through mechanical methods such as ultrasonication, enzymatic digestion, or transposase-mediated fragmentation, each of which influences fragment size distribution and the representation of GC-rich or repetitive regions. Following fragmentation, sequencing adapters are ligated to both ends of the DNA or cDNA fragments to enable immobilization on the flow cell and compatibility with sequencing chemistry. PCR-free library preparation [19] is preferred for high-input DNA samples, such as whole-genome sequencing, to minimize amplification bias. In contrast, low-input or degraded samples, including formalin-fixed paraffin-embedded (FFPE) tissues or single-cell RNA, often require amplification, which can introduce sequence-specific biases such as the overrepresentation of GC- or AT-rich regions and the generation of PCR duplicates.

Depending on study objectives, enrichment strategies are incorporated to target specific genomic regions. Hybrid capture and PCR-based amplicon methods are widely used for DNA libraries, whereas ribosomal RNA depletion or poly(A) selection is commonly employed for RNA-seq applications. The efficiency and specificity of enrichment strategies strongly influence sequencing depth, uniformity, and sensitivity for detecting genetic variants or transcripts. To allow the simultaneous sequencing of multiple samples in a single run, unique barcodes or indices—short sequences of 6–12 base pairs—are incorporated during library preparation. These indices may be added during PCR or adaptor ligation, with unique dual indices and unique molecular identifiers (UMIs) increasingly adopted to minimize index hopping and PCR artifacts, thereby improving data accuracy.

Sequencing technologies broadly employ either short-read or long-read approaches. Short-read sequencing generates reads ranging from 50 to 600 base pairs, in single-end or paired-end formats, and is highly effective for applications such as whole-genome [20-23] exome sequencing [24-27], RNA-seq [28-31] and targeted panels [32-36]. However, its limitations include difficulty in resolving repetitive elements, phasing haplotypes, and detecting structural variants. Long-read sequencing, in contrast, produces reads ranging from kilobases to megabases in length, enabling the assembly of complex genomes, the identification of structural rearrangements, isoform-level transcriptome profiling, and the direct detection of DNA and RNA modifications. Although historically associated with higher error rates, recent advances such as PacBio HiFi circular consensus sequencing and improved base-calling algorithms for Oxford Nanopore Technologies have substantially narrowed the accuracy gap relative to short-read platforms [3, 13, 14, 37].

Technological advances in microfluidics and barcoding strategies have also enabled single-cell sequencing, allowing for the profiling of genomes, transcriptomes, and epigenomes at the cellular level. These methods require highly sensitive library preparation strategies that can amplify picogram-level inputs while minimizing technical noise. More recently, spatial sequencing technologies have emerged, integrating molecular readouts with tissue architecture to preserve spatial context and extend insights beyond dissociated cell populations.

Despite substantial progress, challenges remain in both library preparation and sequencing. Batch effects, sequence-specific biases, incomplete coverage of GC-rich or repetitive regions, and cross-contamination of indices continue to impact data quality. Future innovations

are expected to focus on minimizing input requirements, reducing amplification biases, and enhancing the ability to sequence native nucleic acids, including direct detection of epigenetic modifications. Furthermore, integrative workflows that combine DNA, RNA, chromatin, and protein information within a single library preparation are emerging as powerful approaches for generating comprehensive multi-omic profiles.

Together, advances in library preparation and sequencing technologies continue to expand the scope of NGS applications. Their integration into seed research, quality control, and biotechnology promises to accelerate discovery and enhance the reproducibility, resolution, and translational impact of genomic science.

Next-generation sequencing technology for crop improvement and quality seed production

High-quality seed remains central to addressing global food security. Challenges driven by climate change, limited resources, and population growth requires higher production driven by better seed. Conventional breeding strategies, provide reliable results but suffer from low throughput, high labor demands, and limited scalability, rendering them unsuitable for modern large-scale genomic studies.

Molecular marker-based breeding offers an effective alternative for accelerating crop development and enhancing yield and quality. Key marker types-single nucleotide polymorphisms (SNPs), insertions/deletions (indels), simple sequence repeats (SSRs), and structural variants (SVs)-form the foundation of modern plant breeding programs [38]. Their identification and characterization are closely linked to the rapid evolution of sequencing platforms, particularly next-generation sequencing (NGS).

NGS technologies enable large-scale transcriptome and whole-genome Sequencing across diverse plant taxa, generating comprehensive datasets of genetic variation and gene expression. These resources support the development of functional markers associated with specific agronomic traits. Coupled with advanced bioinformatics, NGS facilitates the efficient discovery of markers and enhances the application of marker-assisted selection (MAS) in breeding pipelines [39].

The convergence of NGS and molecular marker development has transformed plant breeding. Researchers can now uncover novel genetic variants and establish precise markers that were previously inaccessible to traditional methods. These markers are now utilized in large-scale breeding programs for traits such as disease resistance, drought tolerance, and yield stability. For example, research has demonstrated the utility of NGS for identifying SNP-based markers linked to seed traits, accelerating marker-assisted breeding pipelines. NGS has also been utilized to identify genetic variants associated with stress tolerance in barley, based on early seedling traits, highlighting the broad applicability of NGS in genotyping for seed trait improvement [40].

NGS-driven crop improvement and seed quality assurance provided a revolutionary change in agriculture. NGS is now being used extensively for crop improvement, see quality assessment and seed health determination. SNP-based genotyping powered by NGS provides robust hybrid purity testing in crops such as maize and cotton, outperforming earlier SSR- or protein-based assays. It ensures yield stability and strengthens the commercialization of hybrid seeds [40].

Maintaining varietal purity is crucial for the credibility of the seed industry. NGS-based DNA fingerprinting surpasses traditional SSR markers by enabling high-resolution discrimination between varieties. discussed the integration of NGS with automated image-based platforms for seed quality assessment, reinforcing its value in purity testing systems.

Seed-borne pathogens pose a significant threat to crop health and global trade. Metagenomic Sequencing offers unbiased detection of both known and novel pathogens [41] validated an NGS-based viral safety assessment method, illustrating its application in pathogen surveillance for seed lots. Such analysis can be performed in the field using 16S sequencing or metagenome sequencing with Nanopore technology. Short-read Sequencing and other platforms also enable the Sequencing and identification of bacterial, viral, and fungal species in seed samples using NGS-based sequencing approaches. NGS-based analysis of the microbiome enables us to overcome issues related to microbial culture and the time-consuming process of culturing. These methods help ensure compliance with phytosanitary regulations and safeguard international seed exchange.

During repeated seed multiplication cycles, subtle genetic shifts can erode varietal integrity. Whole-genome resequencing helps monitor this drift across generations.

demonstrated how sequencing and imaging tools can be used jointly to detect variation in seed lots, providing breeders with actionable information to maintain genetic consistency [42].

NGS and other genomic approaches can be crucial for breeding crops with higher yields and improved tolerance to biotic and abiotic stresses. The vast untapped genetic diversity, which can be utilized to identify such alleles that contribute to enhanced plant performance, can be leveraged using these novel genomic techniques. Improving crops for complex genetic traits, such as yield, techniques like genomic selection can become extremely handy and an extremely vital tool for the complete dissection of such complex traits. The NGS-based genotyping has increased genomic-estimated breeding value prediction accuracies over other established marker platforms in cereals and other crop species. The benefits of NGS came at a cost-both the infrastructure needed to carry out NGS and the process of collecting sequencing data were prohibitively expensive. However, due to the involvement of vast amounts of costs and being technically challenging, it still cannot be employed in large-scale breeding programs, and MAS remains the choice of breeders for introgression of QTLs. The recent decrease in sequencing costs has enabled the large-scale sequencing of plant genomes for the identification of genetic diversity [38-41, 43-45].

Traditional map-based cloning approaches required ~3 years to identify genes underlying simple Mendelian traits. Similarly, eliminating deleterious alleles from contaminated seed stocks required multi-year selection cycles. Conventional markers, such as SSRs, lacked the resolution to detect single-nucleotide variants. Although gene isolation technologies have accelerated crop improvement, their integration with breeder seed maintenance has remained limited. Whole-genome Sequencing (WGS) enables the rapid identification of causative alleles by detecting variants that show complete linkage with mutant traits. In case studies, NGS reduced timelines for gene discovery from ~3 years to ~1 year, enabling prompt removal of deleterious alleles from seed populations. For example, identification of the swl1-R332P allele facilitated the elimination of a recessive albino phenotype in Hinohikari rice, restoring healthy breeder seeds. However, additional background mutations (~216 polymorphisms) highlighted risks of mutation accumulation over successive generations. Whole-genome sequencing approaches enable the rapid detection of spontaneous mutations, the precise removal of alleles within one year [46] and increased confidence in seed supply chain integrity. To overcome cost hurdles, low-pass Sequencing, which provides sparse but genome-wide coverage, is a viable strategy. Although individual read reliability is low, linkage-based imputation generates accurate genotypes, making low-pass sequencing a practical tool for large-scale MAS. Hence, low-pass Sequencing with imputation provides a viable option for scalable marker-assisted selection and diversity analysis. In contrast, whole-genome Sequencing will help to generate a reference genome and genome-wide associations.

High-quality reference assemblies of crop plants have accelerated evolutionary and comparative genomics. Next-generation sequencing (NGS) enables genome-wide detection of variation, including simple sequence repeats (SSRs), single-nucleotide polymorphisms (SNPs), insertions/deletions (InDels), translocations, fusions, and copy number variations (CNVs). These variations can be converted into genetic markers, particularly SSRs and SNPs. Re-sequencing studies generally follow three approaches:

1. Genome-wide identification of structural variations [43].
2. Population genomics: genotyping-by-sequencing (GBS) and genome-wide association studies (GWAS)[40]
3. Pan-genome analysis [47].

GWAS links genome-wide SNPs to phenotypic traits, initially using microarrays and later NGS. This approach leverages linkage disequilibrium across diverse accessions, genotyped once but phenotyped repeatedly.

A pan-genome represents the total set of genes within a clade (species or genus) and is subdivided into core genes (present in all accessions) and dispensable genes (subgroup- or individual-specific). First described in bacteria (*Streptococcus agalactiae*) in 2005, pan-genomics revealed that ~80% of genes form the core, while ~20% are dispensable.

Although initially limited by sequencing costs and assumptions of low variation in higher organisms, pan-genome analysis has since been applied to numerous crops, including Brassica, maize, wheat, rice, soybean, and pigeon pea [47]. Early studies demonstrated that 15-40% of the plant gene repertoire is dispensable, with

many of these genes linked to biotic and abiotic stress tolerance. Pan-genome captures genetic diversity, which is otherwise missed by a single reference genome. It provides new sources of allelic variation for marker development and genomics-assisted breeding and identifies dispensable genes linked to stress tolerance, flowering, disease resistance, and yield [47].

The first plant pan-genome study sequenced seven wild soybean (*Glycine soja*) accessions [48] identifying dispensable genes related to biomass, flowering, maturity, seed composition, organ size, and disease resistance. Iterative mapping and assembly approaches, employed in *Brassica oleracea* (10 genotypes) [49] wheat (18 genotypes) [50] and rapeseed (53 genotypes) [51] uncovered extensive presence-absence variation (PAV) affecting agronomic traits such as stress tolerance, flowering, and metabolite biosynthesis. Pangenome-based studies in soybean [52, 53] and pigeon pea [54] revealed new genes associated with seed composition and seed weight, respectively.

In Sesame a comparative analysis of old and modern cultivars identified dispensable genes linked to domestication traits. In *Brachypodium distachyon* [55] sequencing of 54 individuals discovered 7,135 novel genes absent from the reference genome, highlighting population-specific gene sets.

In summary, next-generation sequences offer an opportunity for

1. Molecular marker discovery & genotyping to enable SNP/indel-based markers for seed traits.
2. Hybrid seed authentication to confirm parental line identity and hybrid purity.
3. Genome-wide association studies (GWAS) link genomic variants to agronomic seed traits.
4. Transcriptomics (RNA-seq) to reveal expression networks in seed development, dormancy, and stress.
5. An epigenomic study to understand the regulatory mechanisms underlying vigor and longevity.
6. Metagenomics to profile seed-associated microbiomes, critical for plant health.
7. Pangenomics to capture genomic diversity within crop species, identifying novel alleles linked to seed traits.
8. Varietal purity testing using NGS-based DNA fingerprinting offers greater resolution than SSRs.
9. Pathogen detection by metagenomic Sequencing identifies both known and novel seed-borne pathogens.
10. Traceability & certification using DNA barcoding and Sequencing improve transparency in seed supply chains.
11. Monitoring genetic drift through whole-genome resequencing tracks variation across seed multiplication cycles.

NGS technologies—from low-pass sequencing to pangenomics-are reshaping seed research, breeding, and quality control. By enabling rapid gene discovery, scalable MAS, and microbial detection, NGS bridges molecular genetics with practical seed management. Continuous reductions in sequencing costs will further democratize its application, ensuring genetic diversity, varietal purity, and global seed security.

REFERENCES

1. BARBA M, H CZOSNEK AND A HADIDI (2014). Historical perspective, development and applications of next generation sequencing. *Plant Virus*, **8**: 115-134.
2. HALLMANN A (2014). DNA sequencing methods-historical review.
3. PAREEK CS, R SMOCZYNSKI AND A TRETYN (2011). Sequencing technologies and genome sequencing. *Journal of Applied Genetics*, **52(4)**: 413-435.
4. SATAM H (2023). Next-Generation Sequencing Technology: Current Trends and Advancements. *Biology (Basel)*, **12(7)**.
5. WARBURTON PE AND RP SEBRA (2023). Long-Read DNA Sequencing: Recent Advances and Remaining Challenges. *Annu Rev Genomics Hum Genet*, **24**: 109-132.
6. SANGER F, S NICKLEN AND AR COULSON (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences*, **74**: 5463-5467.
7. LI J, X WANG AND Z CHEN (2016). An NGS Workflow Blueprint for DNA Sequencing Data and Analyses. *Genomics, Proteomics & Bioinformatics*, **14**: 11-24.
8. PEREIRA R, J OLIVEIRA AND M SOUSA (2017). Practical Overview of NGS Workflows: from Sample to Variant Interpretation. *Human Genetics*, 136: 1043–1060.
9. SLATKO BE, AF GARDNER AND FM AUSUBEL (2018). Overview of Next-Generation Sequencing Technologies. *Current Protocols in Molecular Biology*, **122(1)**: e59.
10. DEAMER D, M AKESON AND D BRANTON (2016). Three decades of nanopore sequencing. *Nature Biotechnology*, **34**: 518-524.
11. JAIN M (2016). The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. *Genome Biology*, **17**: 239.

12. MCCARTHY A (2010). Third generation DNA sequencing: Pacific Biosciences' single molecule real time technology. *Chemistry & Biology*, **17**: 675-676.
13. RHOADS A AND KF AU (2015). PacBio sequencing and its applications. *Genomics, Proteomics & Bioinformatics*, **13(5)**: 278-289.
14. VAN DIJK EL, et al. (2018). The third revolution in sequencing technology. *Trends in Genetics*, **34(9)**: 666-681.
15. BAYM M (2015). Inexpensive multiplexed library preparation for megabase-sized genomes. *PLoS ONE*, 10(5): e0128036.
16. FROMMER M (1992). A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. *Proceedings of the National Academy of Sciences of the United States of America*, **89**: 1827-1831.
17. HEAD SR (2014). Library construction for next-generation sequencing: Overviews and challenges. *Biotechniques*, **56(2)**: 61-77.
18. ILLUMINA I (2020). DNA Library Prep for Next-Generation Sequencing. *Illumina Technical Note*.
19. KOZAREWA I AND DJ TURNER (2011). Amplification-free library preparation for paired-end Illumina sequencing. *Genome Biology*, **12(2)**: R6.
20. BENTLEY DR (2008). Accurate whole human genome sequencing using reversible terminator chemistry. *Nature*, **456(7218)**: 53-59.
21. LEVY S, G SUTTON, AND PC NG (2007). The diploid genome sequence of an individual human. *PLoS Biology*, **5(10)**: e254.
22. MARDIS ER (2013). Next-generation sequencing platforms. *Annual Review of Analytical Chemistry*, **6**: 287-303.
23. NIELSEN R (2011). Genotype and SNP calling from next-generation sequencing data. *Nature Reviews Genetics*, **12(6)**: 443-451.
24. BAMSHAD MJ, SB NG AND AW BIGHAM (2011). Exome sequencing as a tool for Mendelian disease gene discovery. *Nature Reviews Genetics*, **12(11)**: 745-755.
25. BELKADI A, V PEDERGNANA AND A COBAT (2015). Whole-genome sequencing is more powerful than whole-exome sequencing for detecting exome variants. *Proceedings of the National Academy of Sciences USA*, **112(17)**: 5473-5478.
26. NG SB, KJ BUCKINGHAM, AND C LEE (2010). Exome sequencing identifies the cause of a Mendelian disorder. *Nature Genetics*, **42(1)**: 30-35.
27. NG SB, EH TURNER, AND PD ROBERTSON (2009). Targeted capture and massively parallel sequencing of 12 human exomes. *Nature*, **461(7261)**: 272-276.
28. CONESA A, P MADRIGA AND S TARAZONA (2016). A survey of best practices for RNA-seq data analysis. *Genome Biology*, **17**: 13.
29. MORTAZAVI A (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature Methods*, **5(7)**: 621-628.
30. OZSOLAK F AND PM MILOS (2011). RNA sequencing: advances, challenges and opportunities. *Nature Reviews Genetics*, **12(2)**: 87-98.
31. WANG Z, M GERSTEIN AND M SNYDER (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics*, **10(1)**: 57-63.
32. CHOI M, UI SCHOLL AND W JI (2009). Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. *Proceedings of the National Academy of Sciences USA*, **106(45)**: 19096-19101.
33. GNIRKE A, A MELNIKOV AND J MAGUIRE (2009). Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. *Nature Biotechnology*, **27(2)**: 182-189.
34. HODGES E, Z XUAN AND V BALIJA (2007). Genome-wide in situ exon capture for selective resequencing. *Nature Genetics*, **39(12)**: 1522-1527.
35. MAMANOVA L, AJ COFFEY AND CE SCOTT (2010). Target-enrichment strategies for next-generation sequencing. *Nature Methods*, **7(2)**: 111-118.
36. REHM HL, JS BERG AND LD BROOKS (2015). ClinGen — The Clinical Genome Resource. *New England Journal of Medicine*, **372**: 2235-2242.
37. GOODWIN S, JD MCPHERSON AND WR MCCOMBIE (2016). Coming of age: ten years of next-generation sequencing technologies. *Nature Reviews Genetics*, **17(6)**: 333-351.
38. ATUL S (2024). Molecular Marker-Assisted Selection in Plant Breeding. In: *Plant Molecular Breeding in Genomics Era*. Springer.
39. ADNAN Y (2020). Molecular Markers Improve Abiotic Stress Tolerance in Crops: A Review. *Plants*, **9(10)**.
40. JIANGFENG H (2014). Genotyping-by-sequencing (GBS), an ultimate marker-assisted selection (MAS) tool to accelerate plant breeding. *Frontiers in Plant Science*, 5. 484. doi: 10.3389/fpls.2014.00484
41. ALJUBOORI MN (2024). Application of next-generation sequencing to identify different pathogens. *Frontiers in Microbiology*, **14**.
42. SANTHY V (2023). Molecular Techniques for Testing Genetic Purity and Seed Health. In: *Seed Science and Technology*. Springer.
43. KANDIAH P (2023). Molecular Characterization and Mapping of Stress Resistance Genes Using SNP Platform in Legumes. In: *Legumes: Physiology and Molecular Biology of Abiotic Stress Tolerance*. Springer.
44. QIAN Y (2018). Development and Applications of a High Throughput Genotyping Tool for Polyploid Crops: Single Nucleotide Polymorphism (SNP) Array. *Frontiers in Plant Science*, **9**.
45. RAHUL K (2024). Advances in genomic tools for plant breeding: harnessing DNA molecular markers, genomic selection, and genome editing. *Biological Research*, **57**: 80.
46. BALIMPONYA EG et al. (2022). Seed management using NGS technology to rapidly eliminate a deleterious allele from rice breeder seeds. *Breeding Science*, **72(5)**: 362-371.
47. SCHREIBER M et al. (2024). Plant pangenomes for crop improvement, biodiversity and evolution. *Nature Reviews Genetics*, **25(8)**: 563-577.
48. LI YH (2014). De novo assembly of soybean wild relatives for pan-genome analysis of diversity and agronomic traits. *Nature Biotechnology*, **32(10)**: 1045-1052.
49. BAYER PE (2019). Variation in abundance of predicted resistance genes in the *Brassica oleracea* pangenome. *Plant Biotechnology Journal*, **17(4)**: 789-800.
50. WALKOWIAK S et al. (2020). Multiple wheat genomes reveal global variation in modern breeding. *Nature*, **588(7837)**: 277-283.

51. SONG JM (2020). Eight high-quality genomes reveal pan-genome architecture and ecotype differentiation of *Brassica napus*. *Nature Plants*, **6(1)**: 34-45.

52. LIU Y (2020). Pan-Genome of Wild and Cultivated Soybeans. *Cell*, **182(1)**: 162-176.

53. ZHUANG Y (2022). Phylogenomics of the genus *Glycine* sheds light on polyploid evolution and life-strategy transition. *Nature Plants*, **8(3)**: 233-244.

54. ZHAO J (2020). Trait associations in the pangenome of pigeon pea (*Cajanus cajan*). *Plant Biotechnology Journal*, **18(9)**: 1946-1954.

55. GORDON SP (2017). Extensive gene content variation in the *Brachypodium distachyon* pan-genome correlates with population structure. *Nature Communications*, **8(1)**: 2184.