Journal of Applied Biosciences and Agri-Food Technology 2025, Vol 1(1), 1-9
ISSN:

JABAT
Food Security Faculty
State University of Surabaya

Review Article

Whole-Genome Sequencing Approaches in Vegetable Crops

Debora Marpaung*

Department of Food Science and Technology, Faculty of Agricultural Technology, Brawijaya University, Malang, 65145, Indonesia

Article Info

Article History
Received:
Revised:
Accepted:
Available online:

* Email (Author Corresponding) : deboramarpaung@student.ub.ac.id

ABSTRACT

Whole-genome sequencing (WGS) technologies have fundamentally transformed vegetable crop research, providing essential genomic information that drives innovation in food biotechnology. This review synthesizes current WGS platforms and analytical approaches, explains their evolution, and evaluates their strengths and limitations for vegetable genomics. We trace progress from short-read technologies that enabled the first reference genomes to recent long-read and telomere-to-telomere assemblies that deliver near-complete genomic landscapes. By linking these technical advances to applied outcomes, such as precise trait discovery, accelerated breeding, improved nutritional quality, and enhanced resilience to biotic and abiotic stresses, this review clarifies how WGS serves as a practical toolbox for developing safer, more sustainable, and higher-value vegetable-based foods.

Keywords: Whole Genome Sequencing, Plant Genomics, Pangenome-Based WGS

Introduction

Vegetable crops constitute a critical component of global food systems and human diets. It serves as a primary source of essential vitamins, minerals, dietary fiber, and a diverse array of phytonutrients and functional metabolites (Tripathy et al., 2021). Despite their nutritional significance, many vegetable species face persistent challenges, including genetic erosion, susceptibility to biotic and abiotic stresses, and post-harvest losses, which threaten yield stability and nutritional security (El-Ramady, 2013; Solankey et al., 2021). Recent advances in WGS have driven substantial innovation in our understanding of their genetic architecture, enabling targeted improvement through food biotechnological approaches.

WGS has emerged as an innovative technology in vegetable agriculture, enabling the comprehensive analysis of genetic variations that underpin traits such as yield, disease resistance, and nutritional quality (Liu, 2024; Pandey et al., 2025). By sequencing the entire DNA of a vegetable genome, including coding and non-coding regions, WGS facilitates precision breeding, stress tolerance enhancement, and biodiversity conservation, addressing global challenges like food security and climate change.

Historical Roots of Whole Genome Sequencing

The foundation of WGS relies on pivotal discoveries in molecular biology and sequencing technologies. The discovery of DNA structure in 1953 by James Watson and Francis Crick elucidated DNA’s double-helix structure, revealing how genetic information is stored and replicated (Small, 2023). This provided the basis for understanding genomes as sequences of nucleotide bases. In 1977, Frederick Sanger developed the chain-termination method (Eren et al., 2022). This is the first practical DNA sequencing technique, which

Received on 24 December 2025, Revised on 26 December 2025, Accepted on 29 December 2025

sequenced small DNA fragments with high accuracy, becoming the foundation for DNA sequencing, enabling early insights into plant genetics. However, its limitations in speed, cost, and scalability restricted its application to small genomic regions, making WGS of complex vegetable genomes impractical. Sanger sequencing processes one DNA fragment at a time, requiring manual handling and producing only a few hundred base pairs per reaction (Saini et al., 2023).

The advent of next-generation sequencing (NGS) in the mid-2000s overcame these barriers, revolutionizing genomics by enabling high-throughput, cost-effective sequencing of entire genomes. NGS was introduced around 2005 with platforms like 454 and Illumina enabling massively parallel sequencing, producing millions of short reads (50–300 bp) simultaneously generating gigabases of data per run (A. Kumar et al., 2022). However, its short reads struggled with repetitive and polyploid genomes. Therefore, Third-generation sequencing (TGS) emerged around 2009, addressing these issues by producing long reads (10 kb to >2 Mb), improving the assembly of complex vegetable genomes (Oliveira et al., 2024). TGS focuses on sequencing single DNA molecules in real-time, producing ultra-long reads that span thousands to hundreds of thousands of base pairs, resolving structural variants, and handling polyploid genomes common in vegetables. TGS encompasses two primary platforms, which are Pacific Biosciences (PacBio) Single-Molecule Real-Time (SMRT) Sequencing and Oxford Nanopore Technologies (ONT) Nanopore Sequencing.

Whole Genome Sequencing Methods

WGS methods encompass a diverse set of technologies designed to generate comprehensive DNA sequence information across entire genomes. Over the past two decades, these approaches have evolved from labor-intensive, low-throughput platforms to highly automated systems capable of producing massive datasets with increasing accuracy and read length. Understanding the principles, strengths, and limitations of various WGS methods is crucial for selecting suitable strategies in vegetable genomics and food biotechnology research. This subsection outlines the major sequencing platforms and workflows, highlighting how each contributes to the quality of genome assembly, functional annotation, and downstream biological interpretation.

De Novo WGS

De novo WGS is indispensable for the genomic characterization of unexplored or non-model vegetable species lacking prior reference sequences. This approach entails the fragmentation of genomic DNA, followed by high-throughput sequencing technologies to generate long reads that span repetitive or structurally complex regions and subsequent de novo assembly using algorithms such as de Bruijn graph-based or overlap-layout-consensus (OLC) frameworks (De Luca et al., 2025).

Reference-Based WGS (Resequencing)

Reference-based WGS, also known as resequencing, involves sequencing the genome of a target accession and aligning the resulting reads to a pre-existing reference genome to identify genetic variants, such as single-nucleotide polymorphisms (SNPs), insertions, deletions (indels), and structural rearrangements (Kumawat et al., 2022). This comparative approach offers a rapid and cost-efficient alternative to de novo sequencing, as it leverages the genomic framework of well-characterized model species.

Low-Pass WGS

Low-pass WGS represents a cost-effective genotyping strategy that employs low sequencing coverage ranging from 1× to 5×, typically across large populations (Sogbe et al., 2025). The approach relies on statistical imputation algorithms to infer unobserved genotypes based on linkage disequilibrium patterns and haplotypes derived from a high-quality reference panel. Predominantly implemented using NGS platforms, low-pass WGS enables the efficient capture of genome-wide variation at substantially reduced costs, making it particularly suitable for genomic selection (GS) and association mapping in plant breeding programs.

Pangenome-Based WGS

Pangenome-based WGS represents an advanced genomic framework that captures the full complement of genetic diversity within a species by sequencing multiple individuals or accessions (Taylor et al., 2024). This approach distinguishes core genes that are conserved across all accessions from dispensable or variable genes, which contribute to phenotypic diversity and environmental adaptation. The construction of a pangenome typically integrates de novo assemblies, frequently utilizing third-generation sequencing (TGS) technologies to achieve high contiguity, with reference-based resequencing of broader populations using NGS data. This hybrid strategy allows for comprehensive detection of structural variants, presence–absence variations (PAVs), and novel genomic segments that are often absent in single-reference genomes.

Whole Genome Sequencing (WGS) Prospect for Food Biotechnology

The declining cost and increasing accessibility of high-throughput sequencing technologies have catalyzed an explosion in the number of sequenced vegetable genomes. **Table 1** provides a curated summary of representative vegetable crops that have been sequenced. These foundational genomes are the bedrock upon which modern food biotechnology is built. Whole-genome sequencing, particularly datasets derived from NGS and TGS platforms, has enabled the systematic identification of candidate genes and informative molecular markers, as cataloged in the table, which are being directly leveraged for precision breeding. Thus, the prospect outlined by this growing list of sequenced genomes is one of accelerated, knowledge-driven crop improvement, directly contributing to enhanced food security and nutritional quality.

Table 1. WGS of Vegetable Crops and Their Impacts on Food Biotechnology

Scientific Name	**Sequencing Method**	**Key Advantages**	**Findings**	**References**
Solanum lycopersicum (tomato)	NGS (Illumina HiSeq 2000), TGS (PacBio)	Integration of short and long reads enabled resolution of repetitive regions and domestication-related loci	Domestication and understanding genome-scale breeding	(Takei et al., 2021)
Solanum tuberosum (potato)	TGS (PacBio),	Improved assembly contiguity and variant detection	Elucidate polyploid evolution, genome differentiation, and implications for potato breeding	(Hosaka et al., 2025)

Scientific Name	**Sequencing Method**	**Key Advantages**	**Findings**	**References**
		in a polyploid genome		
Capsicum annuum (pepper)	NGS, TGS	Enhanced the detection of structural variants linked to flavor and disease resistance traits	Capsaicin content, bacterial wilt resistance	(Moreno-Contreras et al., 2024)
Solanum melongena (eggplant)	NGS (resequencing)	Cost-effective SNP discovery across diverse accessions for population analyses	Genetic diversity and associated SNPs	(Yu et al., 2024)
Cucumis sativus (cucumber)	TGS (Oxford Nanopore PromethION24), NGS (Illumina HiSeq 2000	Hybrid assemblies improved genome accuracy while enabling population-scale variant analysis	Genome-wide association studies, marker-assisted selection, and genomic prediction models.	(Seiko et al., 2025)
Cucurbita pepo (pumpkin)	NGS (Illumina HiSeq 2000)	Detection of duplication signatures across the genome	Whole-genome duplication event	(Montero-Pau et al., 2018)
Brassica oleracea var. capitata (cabbage)	NGS	Accurate short-read mapping enabled reliable mitochondrial genome reconstruction	Distinct mitochondrial genome	(K. Yang et al., 2018)
Raphanus sativus (radish)	TGS (PacBio, ONT)	Very long reads resolved complex regions underlying flowering-time regulation	Identifies RsMIPS3 as a key regulator of bolting time	(F. Yang et al., 2025)
Allium cepa (onion)	NGS (Illumina NovaSeq 6000), TGS (PacBio)	Improved the assembly of a large, repetitive genome and supported functional annotation	Molecular breeding, evolutionary studies, and functional genomics	(Cho et al., 2025)
Allium sativum (garlic)	NGS (Illumina NovaSeq 6000)	Efficient variant detection across accessions for	Novel loci and candidate genes	(Y. Wang et al., 2025)

Scientific Name	Sequencing Method	Key Advantages	Findings	References
		candidate-gene discovery		
Apium graveolens (celery)	TGS (PacBio)	Long reads enabled identification of resistance-related regions across complex scaffolds	Candidate resistance genes and genomic regions for introgression	(Lee et al., 2025)
Lactuca sativa (lettuce)	NGS,TGS	Combined sequencing resolved structural diversity relevant to species evolution	Evolutionary relationships, potential molecular markers	(Chu et al., 2024)
Spinacia oleracea (spinach)	TGS (PacBio, ONT)	Telomere-to-telomere assemblies captured repeat-rich regions and structural variation	First complete T2T spinach genome	(She et al., 2025)
Beta vulgaris (beet)	NGS	Population-scale sequencing enabled inference of historical gene flow and admixture	Historical gene flow, admixture, and introgression	(Galewski & McGrath, 2020)
Phaseolus vulgaris (green bean)	NGS	Rapid variant discovery facilitated disease-resistance screening and breeding decisions	Virus-free seeds, developing resistant cultivars, and controlling insect vectors	(Astaraki et al., 2025)
Pisum sativum (common pea)	NGS (Re-sequencing)	Resequencing across lines supported identification of alleles linked to phenology	Early-maturing pea lines	(P. Kumar et al., 2025)
Zea mays subsp. Mays (corn)	NGS (Illumina HiSeq 3000)	Dense variant datasets revealed evolutionary history and adaptive signatures	Revealing evolutionary history, adaptive mechanisms, and functional alleles	(Chen et al., 2022)

Scientific Name	Sequencing Method	Key Advantages	Findings	References
Asparagus officinalis (asparagus)	NGS	Fine-scale variant mapping enabled exploration of sex-chromosome evolution	Early stage of sex chromosome evolution	(Li et al., 2014)
Sechium edule (chayote)	NGS (Illumina NovaSeq 6000), TGS (ONT)	Integration of long and short reads clarified repetitive content and genome expansion	Repetitive-rich genome, polyploidization event, significant gene family expansion	(X. Wang et al., 2025)

Conclusion

WGS has fundamentally transformed vegetable crop science, establishing a foundational pillar for modern food biotechnology. As detailed in this review, the diverse array of WGS methodologies provides a versatile toolkit for research and breeding objectives, enabling the identification of genes and markers underlying agronomically important traits such as disease resistance, stress tolerance, nutritional quality, and yield. Beyond serving as a diagnostic resource, WGS now underpins predictive and precision breeding approaches. Looking forward, integrating WGS with genomic selection, artificial intelligence-driven data analytics, and genome editing technologies promises to accelerate the development of resilient, high-yielding, and nutrient-rich vegetable cultivars. Such synergistic applications will be pivotal in addressing global challenges, including food security, climate adaptation, and human health.

Acknowledgement

The authors would like to thank the Faculty of Agricultural Technology, Universitas Brawijaya, for funding the writing of this review article.

Conflict of interest

The authors declare there are no conflicts of interest.

References

Astaraki, S., Atighi, M. R., & Shams-bakhsh, M. (2025). High-throughput sequencing revealed the symptomatic common bean (Phaseolus vulgaris L.) virome in Iran. *Scientific Reports*, *15*(1), 1–14. https://doi.org/10.1038/s41598-025-85281-y

Chen, L., Luo, J., Jin, M., Yang, N., Liu, X., Peng, Y., Li, W., Phillips, A., Cameron, B., Bernal, J. S., Rellán-Álvarez, R., Sawers, R. J. H., Liu, Q., Yin, Y., Ye, X., Yan, J., Zhang, Q., Zhang, X., Wu, S., … Yan, J. (2022). Genome sequencing reveals evidence of adaptive variation in the genus Zea. *Nature Genetics*, *54*(11), 1736–1745. https://doi.org/10.1038/s41588-022-01184-y

Cho, H., Jung, M., Lee, S. J., Park, J. Y., Zoclanclounon, Y. A. B., Kim, C. W., Han, J. W., Kim, J. S., Kim, D. S., Shin, Y., Hwang, Y. J., Lee, T. H., Lee, S. M., Kang, S. H., Won, S. Y., Kim, J. H., Jang, H. Y., Lee, H. E., Lee, E. S., … Ahn, B. O. (2025). Chromosome-level genome assembly and improved annotation of onion genome (Allium cepa L.). *Scientific Data* , *12*(1), 1–9. https://doi.org/10.1038/s41597-025-04635-3

Chu, L., Du, Q., Zuo, S., Liu, G., Wang, H., Liu, G., Zhao, L., & Xu, G. (2024). Assembly and comparative analysis of the complete mitochondrial genome of Vaccinium carlesii Dunn. *Genomics*, *116*(5), 1–16. https://doi.org/10.1016/j.ygeno.2024.110897

De Luca, Y., Cristaudo, A., Cozzolino, S., Scharmann, M., & Cafasso, D. (2025). A reduced-representation sequencing method for identifying sex-specific molecular markers in plants with large genomes. *Plant Biosystems*, *0*(0), 1–10. https://doi.org/10.1080/11263504.2025.2541732

El-Ramady, H. R. (2013). Integrated nutrient management and postharvest of crops. In *Sustainable Agriculture Reviews: Volume 13* (pp. 163–274). Springer.

Eren, K., Taktakoğlu, N., & Pirim, I. (2022). DNA Sequencing Methods: From Past to Present. *Eurasian Journal of Medicine*, *54*, S47–S56. https://doi.org/10.5152/eurasianjmed.2022.22280

Galewski, P., & McGrath, J. M. (2020). Genetic diversity among cultivated beets (Beta vulgaris) assessed via population-based whole genome sequences. *BMC Genomics*, *21*(1), 1–14. https://doi.org/10.1186/s12864-020-6451-1

Hosaka, A. J., Sanetomo, R., & Hosaka, K. (2025). Allotetraploid nature of a wild potato species, Solanum stoloniferum Schlechtd. et Bché., as revealed by whole-genome sequencing. *Plant Journal*, *121*(1), 1–19. https://doi.org/10.1111/tpj.17158

Kumar, A., Choudhury, B., Dayanandan, S., & Khan, M. L. (2022). Molecular Genetics and Genomics Tools in Biodiversity Conservation. *Molecular Genetics and Genomics Tools in Biodiversity Conservation*, *February*, 1–328. https://doi.org/10.1007/978-981-16-6005-4

Kumar, P., Yadav, S., Rani, M., Narang, D., Singla, D., Dhall, R. K., Chhuneja, P., & Sharma, P. (2025). Genomics assisted mapping of earliness in pea (Pisum sativum L.). *Molecular Biology Reports*, *52*(1), 1–10. https://doi.org/10.1007/s11033-025-10506-2

Kumawat, S., Raturi, G., Dhiman, P., Sudhakarn, S., Rajora, N., Thakral, V., Yadav, H., Padalkar, G., Sharma, Y., Rachappanavar, V., & Kumar, M. (2022). Opportunity and Challenges for Whole-Genome Resequencing-based Genotyping in Plants. *Genotyping by Sequencing for Crop Improvement*, *January 2023*, 38–51. https://doi.org/10.1002/9781119745686.ch3

Lee, C., Epstein, L., Kaur, S., Henry, P. M., Postma-Haarsma, A. D., Monroe, J. G., & Van Deynze, A. (2025). A well-annotated genome of Apium graveolens var. dulce cv. Challenger, a celery with resistance to Fusarium oxysporum f. sp. apii race 2. *Plant Journal*, *122*(5), 1–20. https://doi.org/10.1111/tpj.70251

Li, S. F., Gao, W. J., Zhao, X. P., Dong, T. Y., Deng, C. L., & Lu, L. D. (2014). Analysis of transposable elements in the genome of Asparagus officinalis from high coverage sequence data. *PLoS ONE*, *9*(5), 1–8. https://doi.org/10.1371/journal.pone.0097189

Liu, Y. (2024). The Role of Genome-Wide Association Studies (GWAS) in Vegetable Crop Genetic Improvement: From Yield to Nutritional Value. *International Journal of Horticulture*, *14*.

Montero-Pau, J., Blanca, J., Bombarely, A., Ziarsolo, P., Esteras, C., Martí-Gómez, C., Ferriol, M., Gómez, P., Jamilena, M., Mueller, L., Picó, B., & Cañizares, J. (2018). De novo assembly of the zucchini genome reveals a whole-genome duplication associated with the origin of the Cucurbita genus. *Plant Biotechnology Journal*, *16*(6), 1161–1171. https://doi.org/10.1111/pbi.12860

Moreno-Contreras, V. I., Delgado-Gardea, M. C. E., Ramos-Hernández, J. A., Mendez-Tenorio, A., Varela-Rodríguez, H., Sánchez-Ramírez, B., Muñoz-Ramírez, Z. Y., & Infante-Ramírez, R. (2024). Genome-Wide Identification and Characterization of SNPs and InDels of Capsicum annuum var. glabriusculum from Mexico Based on Whole Genome Sequencing. *Plants*, *13*(22), 1–15. https://doi.org/10.3390/plants13223248

Oliveira, M., Marszałek, K., Kowalski, M., Frolova, A., Łabaj, P. P., Branicki, W., Madureira-Carvalho, Á., da Silva, D. D., & Dinis-Oliveira, R. J. (2024). Sequencing Technologies in Forensic Microbiology: Current Trends and Advancements. *Forensic Sciences*, *4*(4), 523–545. https://doi.org/10.3390/forensicsci4040035

Pandey, A., Unnikrishnan, N., Aparna, R. P. N., & Gayen, D. (2025). Introduction to Omics and Its Application in Crop Improvement. In *From Gene Discovery to Climate-resilient Crops: Omics in Crop Improvement* (pp. 96–115). CABI GB.

Saini, M. K., Gaurav, Harish BM, Kumar, J., & Sanu, K. (2023). DNA Sequencing techniques: Sanger to Next Generation Sequencing. *The Science World A Monthly e Magazine*, *3*(09), 2378–2393.

Seiko, T., Muto, C., Shimomura, K., Yano, R., Kawazu, Y., Sugiyama, M., Kato, K., Tomooka, N., & Naito, K. (2025). Chromosome-level assembly of Cucumis sativus cv. 'Tokiwa' as a reference genome of Japanese cucumber. *Breeding Science*, *75*(2), 85–92. https://doi.org/10.1270/jsbbs.24066

She, H., Liu, Z., Xu, Z., Zhang, H., Wu, J., Cheng, F., Wang, X., & Qian, W. (2025). A complete telomere-to-telomere assembly of the Spinacia oleracea genome reveals Y chromosome evolution and centromere landscape. *Plant Communications*, *6*(9). https://doi.org/10.1016/j.xplc.2025.101410

Small, H. (2023). Bayesian history of science: The case of Watson and Crick and the structure of DNA. *Quantitative Science Studies*, *4*(1), 209–228. https://doi.org/10.1162/qss_a_00233

Sogbe, M., Aliseda, D., Sangro, P., de la Torre-Aláez, M., Sangro, B., & Argemi, J. (2025). Prognostic value of circulating tumor DNA in different cancer types detected by ultra-low-pass whole-genome sequencing: a systematic review and patient-level survival data meta-analysis. *Carcinogenesis*, *46*(1). https://doi.org/10.1093/carcin/bgae073

Solankey, S. S., Kumari, M., Akhtar, S., Singh, H. K., & Ray, P. K. (2021). Challenges and opportunities in vegetable production in changing climate: Mitigation and adaptation strategies. *Advances in Research on Vegetable Production Under a Changing Climate Vol. 1*, 13–59.

Takei, H., Shirasawa, K., Kuwabara, K., Toyoda, A., Matsuzawa, Y., Iioka, S., & Ariizumi, T. (2021). De novo genome assembly of two tomato ancestors, Solanum pimpinellifolium and Solanum lycopersicum var. cerasiforme, by long-read sequencing. *DNA Research*, *28*(1), 1–9. https://doi.org/10.1093/dnares/dsaa029

Taylor, D. J., Eizenga, J. M., Li, Q., Das, A., Jenike, K. M., Kenny, E. E., Miga, K. H., Monlong, J., McCoy, R. C., Paten, B., & Schatz, M. C. (2024). Beyond the Human Genome Project: The Age of Complete Human Genome Sequences and Pangenome References. *Annual Review of Genomics and Human Genetics*, *25*(1), 77–104. https://doi.org/10.1146/annurev-genom-021623-081639

Tripathy, B., Rout, S., Mishra, U. N., & Sahoo, G. (2021). Vegetables : A Potential Source of Nutraceuticals. *Annals of R.S.C.B.*, *25*(4), 17921–17941.

Wang, X., Shen, S., Fu, Y., Cao, R., Wei, Y., & Song, X. (2025). High-quality reference genome decoding and population evolution analysis of prickly Sechium edule. *Horticultural Plant Journal*, *11*(2), 827–838. https://doi.org/10.1016/j.hpj.2024.02.007

Wang, Y., Han, W., Wang, T., Jia, C., Liu, J., Fan, X., & Chen, J. (2025). Elucidating the genetic basis of bulb-related traits in garlic (Allium sativum) through genome-wide association study. *International Journal of Biological Macromolecules*, *284*(P2), 137842. https://doi.org/10.1016/j.ijbiomac.2024.137842

Yang, F., Peng, S., Yuan, S., Ran, M., Li, X., Li, Y., Liu, B., Li, M., Kong, C., Yang, X., Pan, G., Yong, X., Ran, K., Kuang, N., Zhang, D., & Lin, H. (2025). A telomere-to-telomere genome assembly of radish (Raphanus sativus L.) provides insights into QTL mapping of bolting traits. *Journal of Genetics and Genomics*, *xxxx*. https://doi.org/10.1016/j.jgg.2025.07.014

Yang, K., Nath, U. K., Biswas, M. K., Kayum, M. A., Yi, G., Lee, J., Yang, T. J., & Nou, I. S. (2018). Whole-genome sequencing of Brassica oleracea var. Capitata reveals new diversity of the mitogenome. *PLoS ONE*, *13*(3), 1–16. https://doi.org/10.1371/journal.pone.0194356

Yu, C., Yang, Q., Li, W., Jiang, Y., Gan, G., Cai, L., Li, X., Li, Z., Li, W., Zou, M., Yang, Y., & Wang, Y. (2024). Development of a 50K SNP array for whole-genome analysis and its application in the genetic localization of eggplant (Solanum melongena L.) fruit shape. *Frontiers in Plant Science*, *15*(November), 1–12. https://doi.org/10.3389/fpls.2024.1492242