



Plant Archives

Journal homepage: <http://www.plantarchives.org>

DOI Url : <https://doi.org/10.51470/PLANTARCHIVES.2026.v26.supplement-2.354>

PANGENOMICS- AN EMERGING FIELD IN CROP IMPROVEMENT: A REVIEW

Anmol Sharma

Department of Genetics and Plant Breeding, CSK Himachal Pradesh Krishi Vishvavidyalaya,
Palampur, HP 176062, India

E-mail: anmolsharmagpb@gmail.com

(Date of Receiving : 20-04-2026; Date of Revision : 23-06-2026; Date of Acceptance : 09-07-2026)

ABSTRACT

Climate change, increase food demand, decreasing resources requires various transformation in approaches in crop breeding. This review gives a brief understanding of pangenomics, how it captures wider variations and applications it unfolds. Pangenomics is one such emerging approach that capture wide range of genetic diversity in plants beyond the single reference genome. However, its application in plant breeding is often limited by the complexity of assembling large polyploid genomes, high computation advancements. Despite these limitations, pangenomics is expected to play a central role in precision breeding and crop improvement. Overall, pangenomics represents a shift in plant genomics offering opportunities to explore genetic diversity and improve future breeding programmes.

Keywords : Climate change, pangenomics, plant breeding, genetic diversity, crop improvement

Introduction

Crop improvement has become important to ensure global food security under the challenges of rapid population growth, climate change and increasing biotic and abiotic stresses. Traditional breeding methods though have focused in improving plant varieties, but these approaches are often time consuming and may not capture full genetic diversity. In light of these challenges, there is a need of integration of modern breeding technologies with the aim to accelerate genetic gain in crop species.

The evolution of high throughput genomic technologies, including Pacific Biosciences High-Fidelity (PacBio HiFi) sequencing, Oxford Nanopore Technologies long-read sequencing, and high-throughput chromosome conformation capture (Hi-C) sequencing, has improved the accuracy of plant genome (Garg *et al.*, 2024). These advances have laid the foundation of the construction of high-resolution plant pangenomes by including the assemblies whether of cultivated and closely wild relatives, capturing core and variable genomic components (Guo *et al.*, 2025; Jayakodi *et al.*, 2024; He *et al.*, 2022).

Genetic variation within gene pool of crop species is the result of number of changes ranging from single

nucleotide polymorphisms (SNPs) to large structural variation (SVs). Understanding the processes underlying variation and exploiting it are critical to maintain and improve the crop productivity (Brozynska *et al.*, 2016). Structural variations include insertions, deletions, duplications and translocation. Unlike SNPs, structural variations (SVs) can better alter gene content and genome architecture, thereby influencing phenotypic traits and adaptive potential (Hu *et al.* 2024). The idea of structural variation is exploited through advance field of plant science, pangenomics. A wave of pangenomics studies in major crops, rice (Wang *et al.*, 2018; Zhao *et al.*, 2018), maize (Hirsch *et al.*, 2014.), soybean (Li *et al.*, 2014), have shed a new light in crop diversity.

What is pangenomics?

The amount of genomic variation occurring within the species cannot be fully established by the traditional single reference genome (Saxena *et al.*, 2014; Ballouz *et al.*, 2019). The reason is that many bioinformatics analyses use a reference genome to compare with new sample. If the reference does not contain genomic sequence similar to sample sequence, reads from the sample will align poorly or will not align at all. The solution to this problem is an approach

that capture full amount of genetic variation in the species as compared to traditional single reference genome, the solution is pangenomics. Pan-genome refers to the full complement of genes of species which can be partitioned into the core genes, shared by all individuals and a set of dispensable genes partially shared by the individuals (Tettelin *et al.*, 2005).

Pangenome studies have focused on presence and absence of genes in individuals. The amount of gene content variation is influenced by changes occurring from SNPs to large structural variations, the more extreme being the presence and absence of genes.

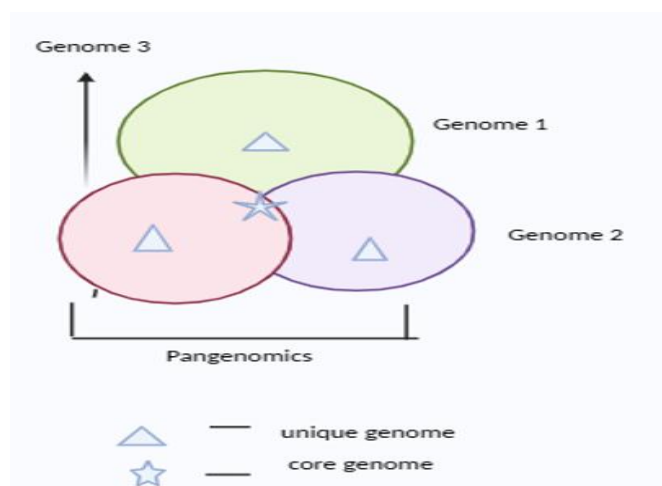


Fig. 1: Three genomes 1, 2 and 3 showing the amount core genome (common to all three), unique genome (possessed by individual species only)

How are Pangenomics generated?

Pangenomics are generated through numerous methodologies including comparative denovo, linear pangenome approach, graph based pangenome. The comparative denovo starts with assembling individual genomes followed by characterizing the identified genes through whole genome comparison. The iterative and map to pan approach starts with map first and assemble later workflow. In this sequencing reads are mapped against chosen reference genome. The reads that fail to align are assembled into novel contigs, which are then introduced in linear pangenomic network. This study is largely useful for large scale population studies as it supports the detection of genomic variants including single nucleotide polymorphisms, short insertion/deletions, presence and absence of variations (PAVs).

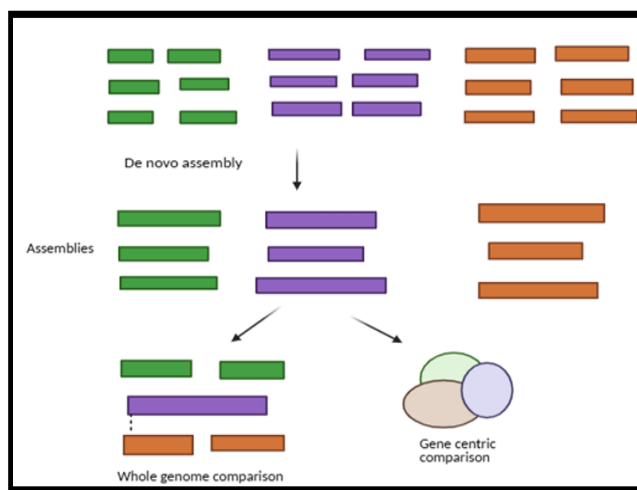


Fig. 2: De novo assembly in a pangenome

In recent years, a graph based pangenome approach have become popular. Instead of representing genome as linear structure, this method uses graph structure to capture genomic information more effectively. Such graph represents useful structural variants, including duplications, deletions, inversions.

Compared to linear reference genome, graph based pangenome represents genomic information of individuals more accurately. However, despite these advantages their use faces difficulty graph construction and difficulty in interpreting graph based genomic structures.

Understanding structural variations as major cause of dispensable genes. The dispensable gene of pangenome are driven by the structural variations (Xu *et al.*, 2012; Mace *et al.*, 2013). Therefore, understanding structural variation in species is important to understand full range of genes present in species for effectively utilizing this genetic diversity in crop improvement programmes.

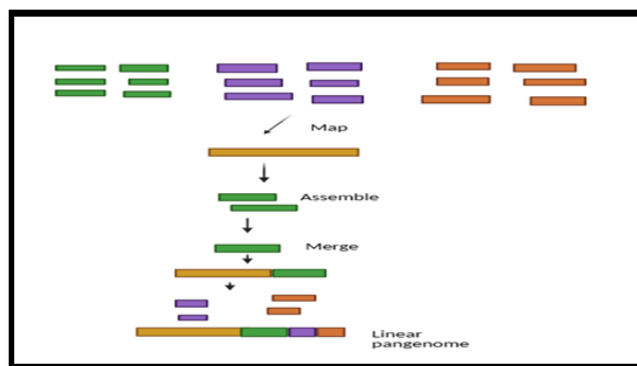


Fig. 3: Linear pangenome where three genomes are mapped, assembled following merge to construct a linear pangenome

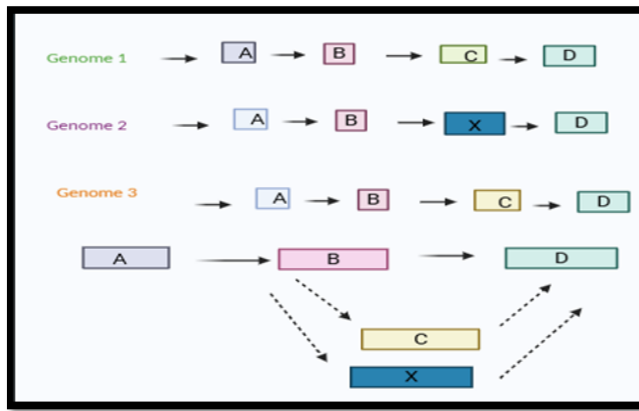


Fig. 4: Graph based pangenome with A, B, C, X, D represents nodes and edges in A→B and the others represent the edge.

Also, A→B→D represents core genome pathway in graphical pangenome, B→X→D represents an accessory pathway

Application of pangenomes

By incorporating multiple genomes, pangenomes revealed genetic diversity that is often absent in the reference genome. Pangenome can be used to map the dispensable genome, which has been increasingly recognized as major contributor to phenotypic

diversity. Recently, pangenomic studies have demonstrated that variable genomic regions are found to be associated with genes controlling key traits such as flowering time, yield components, disease resistance as well as responses to various abiotic stresses (Danilevicz *et al.*, 2020; Aggarwal *et al.*, 2022; Schreiber *et al.*, 2024; Du *et al.*, 2025)

The structural variations identified in pangenomics studies contribute substantially to the variation in gene expression and observable traits (Jeffares *et al.*, 2017) and have been associated with significant agronomic traits. Pangenome can be used to map the important dispensable genes, core genes, accessory genes.

The construction of pangenome characterizes pangenomics associated variation in crop species as well as comparison with exotic and wild relatives. This process can identify genes lost in breeding and selection.

Recent studies have also highlighted the use of pangenomic associated with CRISPR-Cas9 for more effective identification and modification of causal variants, underlying complex agronomic traits (Zhang *et al.*, 2020; Chen *et al.*, 2022; Li *et al.*, 2024; Schreiber *et al.*, 2024; Du *et al.*, 2025).

Table 1: Recent advances in pangenomics

S.No	Crop	Study year	Key findings	Significance	References
1	Rice	Wild cultivated rice pangenome (2025)	Constructed pangenome using 145 genomes; identified ~3.87 Gb novel sequences absent in reference genome	Revealed hidden genetic diversity useful for yield improvement and stress tolerance	Guo <i>et al.</i> , 2025
2	Rice (wild)	Multi species wild cultivated pangenomes (2026)	Identified gene loss (15,000–26,000 genes) during domestication and high PAV (presence–absence variation)	The study helped to understand domestication and stress adaptation (e.g., salt tolerance)	Tao <i>et al.</i> , 2026
3	Maize	Gene family pangenome analysis (2025)	Core, dispensable and private genes were identified in PLD gene family across 26 genomes.	Demonstrated how pangenomes capture variation missed by single reference genomes	Li <i>et al.</i> , 2025
4	Sorghum	Structural variation pangenome (2024; 2026)	Widespread structural variation affecting important agronomic traits were identified.	Important for improving drought resistance and global crop productivity	Zhang <i>et al.</i> , 2024; Morris <i>et al.</i> , 2026
5	Wheat, soybean, chickpea etc.	Comparative pangenome development (2025)	Demonstrated expansion of pangenome size with more genomes; hidden allele was revealed.	Supports breeding for yield, resilience, and adaptation	Ruperao <i>et al.</i> , 2025
6	Orphan crops	Pangenomics in underutilised crops (2025)	Enabled discovery of adaptive and nutritional traits in lesser-studied crops	Important for climate resilience and food security	Hu <i>et al.</i> , 2025

Challenges and opportunities in field of pan genomics

Though pan genomics is powerful approach in capturing genetic diversity, yet it faces some of the

challenges that limits its wider application. Major limitations are high computational requirements for building an analyzing pangenomes involves large scale sequencing data. Accurate pan genome construction

depends on high quality long read genome assembly, costly and still not available for many crop species including orphan crops. In addition, presence of highly repetitive genomic regions makes alignment and annotation difficult.

However, pangenomics offers a wide range of opportunities by capturing comprehensive genetic diversity beyond single reference genome, enabling the identification of structural variants associated with key agronomic traits (Zanini *et al.*, 2021). It enhances molecular breeding and genomic selection, increasing genetic gains by improving marker discovery, trait mapping, also accelerating the development of climate-resilient crops (Aggarwal *et al.*, 2022). Thus, pangenomics has emerged to be a powerful tool to bridge the gap between genomic diversity and practical crop improvement, paving the way towards sustainable and climate resilient agriculture.

Conclusion

The decrease in cost of sequencing and increasing quality have favoured the development and construction of pangenomes of cereal as well as underutilised crops. Pangenomics has shifted our understanding of genetic diversity moving beyond single reference genome to capture the complete genome of the species. By identifying core, accessory and unique genes, pangenomics have helped scientists to identify structural variations, evolutions and agronomically important traits useful in the field of crop improvement. Recent advancement of AI and machine learning have also advanced the pangenomics research enabling the genome annotations, variant calling, trait prediction and regulation studies.

Overall, pangenomics offers a framework to identify the genetic diversity present in the crops to address future agriculture. Its integration into crop improvement will improve breeding strategies and support development of climate resilient agricultural system.

Declarations:

Conflict of interest: The author declare that they have no known competing financial interests or personal relationships that could have appeared to influence this paper.

References

- Aggarwal, S.K., Singh, A., Choudhary, M., Kumar, A., Rakshit, S., Kumar, P., Bohra, A. and Varshney, R.K. (2022). Pangenomics in microbial and crop research: Progress, applications, and perspectives. *Genes*, **13**: 598. doi:10.3390/genes13040598
- Ballouz, S., Dobin, A. and Gillis, J.A. (2019). Is it time to change the reference genome? *Genome Biology*, **20**: 159. doi:10.1186/s13059-019-1774-4
- Brozynska, M., Furtado, A. and Henry, R.J. (2016). Genomics of crop wild relatives: Expanding the gene pool for crop improvement. *Plant Biotechnology Journal*, **14**: 1070–1085. doi:10.1111/pbi.12454
- Chen, Y., Bayer, P.E., Batley, J. and Edwards, D. (2022). Expanding gene-editing potential in crop improvement with pangenomes. *International Journal of Molecular Sciences*, **23**: 2276. doi:10.3390/ijms23042276
- Danilevicz, M.F., Tay Fernandez, C.G., Marsh, J.I., Bayer, P.E. and Edwards, D. (2020). Plant pangenomics: Approaches, applications and advancements. *Current Opinion in Plant Biology*, **54**: 18–25. doi:10.1016/j.pbi.2019.12.005
- Du, Z.Z., He, J.B. and Jiao, W.B. (2025). Plant graph-based pangenomics: Techniques, applications, and challenges. *aBIOTECH*, **6**: 361–376. doi:10.1007/s42994-025-00206-7
- Garg, V., Bohra, A., Mascher, M., Spannagl, M., Xu, X., Bevan, M.W., Bennetzen, J.L. and Varshney, R.K. (2024). Unlocking plant genetics with telomere-to-telomere genome assemblies. *Nature Genetics*, **56**: 1788–1799. doi:10.1038/s41588-024-01830-7
- Guo, D., Li, Y., Lu, H., Zhao, Y., Kurata, N., Wei, X., Wang, A., Wang, Y., Zhan, Q., Fan, D. and Zhou, C. (2025). A pangenome reference of wild and cultivated rice. *Nature*, **642**: 662–671. doi:10.1038/s41586-025-08883-6
- He, Y., Chapman, M.A. and Zhou, M. (2022). Beyond a reference genome: Pangenomes and population genomics of underutilized and orphan crops for future food and nutrition security. *New Phytologist*, **234**: 1583–1597. doi:10.1111/nph.18021
- Hirsch, C.N., Foerster, J.M., Johnson, J.M., Sekhon, R.S., Muttoni, G., Vaillancourt, B., Peñagaricano, F., Lindquist, E., Pedraza, M.A., Barry, K. and de Leon, N. (2014). Insights into the maize pan-genome and pan-transcriptome. *The Plant Cell*, **26**: 121–135. doi:10.1105/tpc.113.119982
- Hu, H., Scheben, A., Wang, J., Li, F., Li, C., Edwards, D. and Zhao, J. (2024). Unravelling inversions: Technological advances, challenges, and potential impact on crop breeding. *Plant Biotechnology Journal*, **22**: 544–554. doi:10.1111/pbi.14224
- Hu, H., Zhao, J., Thomas, W.J.W., Batley, J. and Edwards, D. (2025). The role of pangenomics in orphan crop improvement. *Nature Communications*, **16**: 118. doi:10.1038/s41467-024-55260-4
- Jayakodi, M., Lu, Q., Pidon, H., Rabanus-Wallace, M.T., Bayer, M., Lux, T., Guo, Y., Jaegle, B., Badea, A., Bekele, W. and Brar, G.S. (2024). Structural variation in the pangenome of wild and domesticated barley. *Nature*, **636**: 654–662. doi:10.1038/s41586-024-08187-1
- Jeffares, D.C., Jolly, C., Hoti, M., Speed, D., Shaw, L., Rallis, C., Balloux, F., Dessimoz, C., Bähler, J. and Sedlazeck, F.J. (2017). Transient structural variations have strong effects on quantitative traits and reproductive isolation in fission yeast. *Nature Communications*, **8**: 14061. doi:10.1038/ncomms14061
- Li, S.N., Li, Y.L., Sun, M.H., Sun, Y., Li, X., Cai, Q., Wang, Y. and Zhang, J.G. (2025). Pangenome-wide identification, evolutionary analysis of maize *ZmPLD* gene family, and functional validation of *ZmPLD15* in cold stress tolerance. *Plants*, **14**: 3858. doi:10.3390/plants14243858
- Li, Y.H., Zhou, G., Ma, J., Jiang, W., Jin, L.G. and Zhang, Z., Guo, Y., Zhang, J., Sui, Y., Zheng, L. and Zhang, S.S.

- (2014). *De novo* assembly of soybean wild relatives for pan-genome analysis of diversity and agronomic traits. *Nature Biotechnology*, **32**: 1045–1052. doi:10.1038/nbt.2979
- Mace, E.S., Tai, S., Gilding, E.K., Li, Y., Prentis, P.J., Bian, L., Campbell, B.C., Hu, W., Innes, D.J., Han, X., and Cruickshank, A. (2013). Whole-genome sequencing reveals untapped genetic potential in Africa's indigenous cereal crop sorghum. *Nature Communications*, **4**: 2320. doi:10.1038/ncomms3320
- Morris, G.P., Harder, A.M., Healey, A.L., McLaughlin, C.M., Rifkin, J.L., Cruet-Burgos, C., Jenkins, J.W., Shu, S., Spiekerman, J.J., VanGessel, C.J., Agnew, E. *et al.* (2026). A sorghum pangenome reference improves global crop trait discovery. *Nature*, **652**: 1245–1253. doi:10.1038/s41586-026-10229-9
- Ruperao, P., Rangan, P., Shah, T., Sharma, V., Rathore, A., Mayes, S. and Pandey, M.K. (2025). Developing pangenomes for large and complex plant genomes and their representation formats. *Journal of Advanced Research*, **78**: 29–46. doi: 10.1016/j.jare.2025.01.052
- Saxena, R.K., Edwards, D. and Varshney, R.K. (2014). Structural variations in plant genomes. *Briefings in Functional Genomics*, **13**: 296–307. doi:10.1093/bfpg/elu016
- Schreiber, M., Jayakodi, M., Stein, N. and Mascher, M. (2024). Plant pangenomes for crop improvement, biodiversity and evolution. *Nature Reviews Genetics*, **25**: 563–577. doi:10.1038/s41576-024-00691-4
- Tao, H., Wu, S., Wu, G., Chen, X., Sun, Y., Zhu, X., Chen, L., Feng, C., Kabore, M.A.F., Woldegiorgis, S.T. and Zhang, L. (2026). Pan-genome of wild and cultivated rice uncovers genetic diversity, lost and selected sequences during rice domestication. *Rice*, **19**: 21. doi:10.1186/s12284-026-00893-w
- Tettelin, H., Massignani, V., Cieslewicz, M.J., Donati, C., Medini, D., Ward, N.L., Angiuoli, S.V., Crabtree, J., Jones, A.L. and Durkin, A.S. (2005). Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: Implications for the microbial pan-genome. *Proceedings of the National Academy of Sciences of the United States of America*, **102**: 13950–13955. doi:10.1073/pnas.0506758102
- Wang, W., Mauleon, R., Hu, Z., Chebotarov, D., Tai, S., Wu, Z., Li, M., Zheng, T., Fuentes, R.R. and Zhang, F. *et al.* (2018). Genomic variation in 3,010 diverse accessions of Asian cultivated rice. *Nature*, **557**: 43–49. doi:10.1038/s41586-018-0063-9
- Xu, X., Liu, X., Ge, S., Jensen, J.D., Hu, F., Li, X., Dong, Y., Gutenkunst, R.N., Fang, L. and Huang, L. *et al.* (2012). Resequencing 50 accessions of cultivated and wild rice yields markers for identifying agronomically important genes. *Nature Biotechnology*, **30**: 105–111.
- Zanini, S.F., Bayer, P.E., Wells, R., Snowdon, R.J., Batley, J., Varshney, R.K., Nguyen, H.T., Edwards, D. and Golicz, A.A. (2022). Pangenomics in crop improvement—from coding structural variations to finding regulatory variants with pangenome graphs. *The Plant Genome*, **15**(1): e20177. doi:10.1002/tpg2.20177
- Zhang, Z., Gomes Viana, J.P., Zhang, B., Walden, K.K.O., Müller, P.H., Moose, S.P., Morris, G.P., Daum, C., Barry, K.W., Shakoor, N. and Hudson, M.E. (2024). Major impacts of widespread structural variation on sorghum. *Genome Research*, **34**: 286–299. doi:10.1101/gr.278396.123
- Zhao, Q., Feng, Q., Lu, H., Li, Y., Wang, A., Tian, Q. and Zhan, Q. *et al.* (2018). Pan-genome analysis highlights the extent of genomic variation in cultivated and wild rice. *Nature Genetics*, **50**: 278–284. doi:10.1038/s41588-018-0041-z
- Zhu, H., Li, C. and Gao, C. (2020). Applications of CRISPR–Cas in agriculture and plant biotechnology. *Nature Reviews Molecular Cell Biology*, **21**: 661–677. doi:10.1038/s41580-020-00288-9