Volume 4, Issue 3
March I 2026

Agri Roots
e- Magazine

www.agrirootsmagazine.in

ISSN: 2583-9071

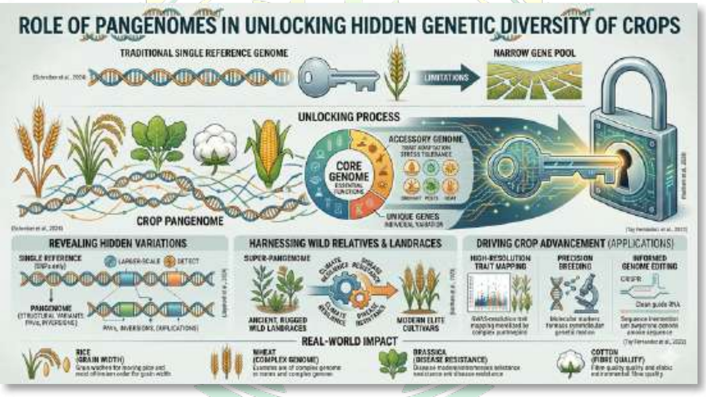
Role of Pangenomes in Unlocking Hidden Genetic Diversity of Crops

ARTICLE ID: 0347

Priyanka Rana

PhD Scholar, Department of Genetics and Plant Breeding, Palli Siksha Bhavana, Visva-Bharati University, Sriniketan, West Bengal – 731236

Global agriculture is now facing unprecedented challenges, including continuous climate change, population growth, and the consequent rise in food demand. These climatic changes affect weather patterns, irregular rainfall, temperature fluctuations, and an increased incidence of extreme climatic events, which lead to stress in crops and overall yield reduction.



An increase in population, along with the predicted impact of climate change, creates a pit in adequate supply, which needs to be filled. To fulfil this increasing food demand and urgency to develop resilient varieties of staple, emerging and orphan crops. Plant genomes are dynamic and environmentally responsive, serving as reservoirs of genetic diversity. However, domestication and intensive breeding have narrowed the genetic base of many cultivated crops, leading to the loss of valuable alleles. Traditionally, crop genomics has relied on a single reference genome (SRG), which represents only one individual and cannot capture the full spectrum of intraspecific variation.

In contrast, a pangenome encompasses the complete set of genes present across multiple individuals of a species. By classifying genes into core, dispensable (accessory), and unique categories, pangenomics provides a more comprehensive genomic framework. This expanded perspective enables the discovery of hidden alleles and structural variations that can enhance crop resilience, productivity, and nutritional quality in the face of climate uncertainty.

Concept of Pangenome

The term "pan-genome" was first used in 2000 in cancer genomics research, and in 2001, G.D. Ehrlich proposed the Distributed Genome Hypothesis, suggesting that not all members of a species share the same set of genes. In 2005, the word "pangenome" gained its modern meaning when used to describe genomic differences among multiple isolates of *Streptococcus agalactiae*. The concept was later extended to plants in 2007, where analysis of maize inbred lines Mo17 and B73 revealed significant genomic differences, highlighting intraspecific variation.

The core genome comprises genes present in all individuals and is essential for fundamental biological functions and species survival. The dispensable (accessory) genome includes genes present in some but not all individuals, contributing to adaptation, stress tolerance, and genetic diversity important for crop improvement.

During domestication and breeding, crops go through a bottleneck, meaning their genetic diversity becomes narrow due to selection and repeated use of a few genotypes. As a result, many useful genes are lost in modern cultivation. Although these dispensable genes are not necessary for the survival of the crop as core genes are, these genes help tolerate stress such as drought, salinity, and diseases. Wild crop relatives still contain this hidden genetic diversity, and pangenome studies help to identify these lost genes.

Presence–absence variation (PAV) represents a major component of genomic diversity captured by pangenomes, where certain genes are present in some individuals but completely absent in others. PAV analysis helps identify dispensable or variable genes that may contribute to important agronomic traits such as stress tolerance, disease resistance, and environmental adaptation. By comparing PAV patterns across wild relatives, landraces, and elite cultivars, researchers can detect genes lost during domestication and breeding. This information provides valuable targets for reintroduction through breeding or genome editing to enhance crop resilience and productivity.

Importance of Hidden Genetic Diversity in Crops

They can be gene-oriented, identifying the core genomes and dispensable/accessory genes. The sequence-oriented pangenome model includes genomic sequence variation, including single-nucleotide variants, insertions, deletions and structural variants. Compared to traditional linear references, pangenomes are less biased and improve mapping accuracy, variant identification, genotyping accuracy and the ability to link genes with phenotypes of interest.

Graph-based pangenomes overcome reference bias by integrating multiple genome assemblies into a non-linear structure where alternative alleles and structural variants are represented as branching paths. This improves the detection of presence–absence variation and complex structural rearrangements that are often missed in single-reference analyses. By enabling more accurate variant calling and allele mining across diverse germplasm, graph-based frameworks strengthen trait discovery and breeding applications, although they require advanced computational tools and resources.

Applications of Pangenomics in Crop Improvement

Pangenomics offers a powerful framework for crop improvement by capturing the full genomic diversity within a species rather than relying on a single reference genome. A major advantage is the identification of genes absent in modern elite cultivars but preserved in landraces and wild relatives, many of which are associated with stress tolerance, disease resistance, and environmental adaptation. Pangenomes can be created for many purposes, using a variety of genomic data formats and methods.

Pangenome analyses enable the detection of structural variations, including insertions, deletions, inversions, duplications, translocations, and copy-number variations. These large-scale genomic changes often influence key agronomic traits yet remain undetected in single-reference analyses. By uncovering such hidden variation, pangenomics broadens the genetic base of breeding populations and reduces genetic vulnerability.

The distinction between core and accessory genomes further clarifies the impact of domestication, as accessory genes are frequently enriched for adaptive and stress-responsive functions. Comparative analysis across diverse germplasm helps identify beneficial alleles that have been lost during crop improvement. Additionally, pangenomes enhance precision breeding by improving marker development, genomic selection accuracy, and candidate gene discovery. Integration with genome editing technologies enables targeted reintroduction of adaptive alleles. Overall, pangenomics reconnects modern cultivars with diverse gene pools, strengthening yield stability, climate resilience, and sustainable crop production.

Selected Case Studies

Pangenome studies across major crops have demonstrated the significance of hidden genetic diversity. To date, there are numerous case studies in the field of crop pangenomics, some involving staple crops and others involving orphan crops. *Brassica rapa*, *Glycine soja*, and *Oryza sativa* were the first three plant species to have their pangenome constructed in 2014. In the case of rice and wheat, a light-shotgun sampling metagenomic approach has been used to create pangenomes.

The rice pangenome was constructed using 1,483 rice lines by assembling only the sequences that did not match the reference genome, grouping them by subspecies. Although this method was cost-effective and allowed analysis of many lines, it could not trace genes back to individual plants and missed rare genes due to low sequencing depth. Even with these limitations, many new non-reference genes were identified. Importantly, a large proportion of metabolic and agronomic traits were linked to these variable, non-reference regions, highlighting the value of pangenomics in crop improvement.

The extensive hexaploid wheat genome faces difficulties for pan-genome development employing a metagenomic methodology. From 18 cultivars, researchers determined that the complete wheat pan-genome comprises 140,500 genes, of which roughly 81,070 are core genes.

Brassica pangenomes were developed using multiple accessions of *B. rapa*, *B. oleracea*, and *B. napus*, revealing thousands of genes beyond the reference genome, with a substantial proportion classified as core and several hundred unique to each species. Earlier iterative assembly approaches in *B. oleracea* also expanded the genome size and uncovered gene presence–absence variation, including flowering-time genes linked to early flowering. Later, a multi-species graph-based "super-pangenome" integrating 41 genomes further improved the representation of structural diversity across Brassica species and led to the development of the Brassica Panache web portal for data access and analysis.

Challenges and Future Perspectives

Despite their immense potential, crop pangenomes face several challenges. The assembly and integration of multiple high-quality genomes into graph-based structures require substantial computational resources and advanced bioinformatics expertise. Accurate detection and interpretation of structural variants and presence–absence variations remain technically demanding, particularly in large, repetitive plant genomes. Another major challenge lies in distinguishing functionally meaningful variation within the accessory genome from neutral diversity. Robust functional validation through transcriptomics, proteomics, and phenomics is essential to translate genomic discoveries into breeding applications.

Future advancements are expected through the integration of long-read sequencing technologies, improved graph-based frameworks, and multi-omics approaches. These developments will enhance allele discovery, functional annotation, and the implementation of precision breeding strategies.

Conclusion

Pangenomics represents a transformative shift in crop genomics by moving beyond single-reference genome models to capture the full spectrum of genetic and structural diversity within species. By uncovering hidden genes, presence–absence variations, and structural rearrangements, it reveals reservoirs of adaptive potential that were previously overlooked in conventional genomic analyses. This comprehensive view of genomic diversity not only deepens our understanding of domestication and evolutionary processes but also provides a powerful foundation for allele mining and precision breeding.

As agriculture faces mounting challenges from climate change, genetic erosion, and increasing global food demand, harnessing the diversity preserved within pangenomes—particularly from diverse germplasm and wild relatives—will be essential for developing resilient, high-yielding, and stress-tolerant crop varieties. With continued technological advancements and integration into breeding pipelines, pangenomics is poised to become a central tool for sustainable crop improvement and long-term food security.

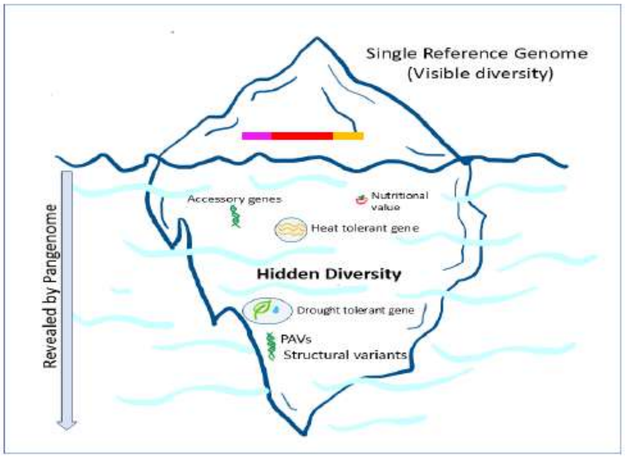


Figure 1. Conceptual representation of hidden genetic diversity revealed by pangenomics

The iceberg model illustrates the limitation of a single reference genome (visible portion above water), which captures only a fraction of the total genetic diversity within a crop species. The larger submerged portion represents hidden genetic variation, including accessory genes, presence–absence variations (PAVs), and structural variants. These hidden components often contain adaptive genes associated with drought tolerance, heat tolerance, disease resistance, and nutritional traits, which are uncovered through pangenome analysis.

References

1. Hu, H., Zhao, J., Thomas, W. J. W., Edwards, D. and Batley, J. (2024). The role of pangenomics in orphan crop improvement. *Nature Communications*. 16:118. https://doi.org/10.1038/s41467-024-55260-4.
2. Kuluev, B. R., Chemeris, D. A., Gerashchenkov, G. A., Kuluev, A. R., Mikhailova, E. V., Vershinina, Z. R., Baymiev, An. Kh., Garafutdinov, R. R., Baymiev, Al. Kh., Chemeris, A. V. (2025). Pangenomics of plants. *Biomics*. 17(1): 42–64. DOI: 10.31301/2221-6197.bmcs.2025-4.
3. Li, L., Goltsman, E., Goodstein, D., Wu, G. A., Rokhsar, D. S. and Vogel, J. P. (2021). Plant Pan-Genomics Comes of Age. *Annual Review of Plant Biology*. 72: 411–435.
4. Matthews, C. A., Watson-Haigh, N. S., Burton, R. A. and Sheppard, A. E. (2024). A gentle introduction to pangenomics. *Briefings in Bioinformatics*, 25(6), bbae588.
5. Petereit, J., Bayer, P. E., Thomas, W. J. W., Tay Fernandez, C. G., Amas, J., Zhang, Y., Batley, J. and Edwards, D. (2022). Pangenomics and Crop Genome Adaptation in a Changing Climate. *Plants*, 11, 1949. https://doi.org/10.3390/plants11151949.