SUNSHINE AGRICULTURE
e-Newsletter

ISSN (E): 2583 – 0821

Sun. Agri.:e- Newsletter, (2026) *6*(10), 153-158

Article ID: 706

Pangenomes and Super-Pan: Why One Reference Genome is No Longer Enough for Breeders

Pranghita Das[1], Parinita Borah[2]
Jyoti I Tondi[3], Aarti Yadav[4]
Nikhil Singh[5]

[1]Ph.D. Research Scholar, Department of Genetics and Plant Breeding
Uttar Banga Krishi Vishwavidyalaya
Pundibari, Coochbehar
[2]Assistant Professor, Department of Plant Breeding and Genetics
Arunachal University of Studies
Namsai, Arunachal Pradesh
[3]Ph.D. Scholar, Department of Genetics and Plant Breeding, University of Agricultural Sciences, Dharwad
[4]M.Sc. Scholar, Department of Agronomy, College of Agriculture
Suwana, Bhilwara, Rajasthan
[5]M.Sc. Scholar, Department of Genetics and Plant Breeding, Acharya Narendra Dev University of Agriculture and Technology, Kumarganj, Ayodhya

Open Access

*Corresponding Author
Pranghita Das*

Available online at
www.sunshineagriculture.vitalbiotech.org

Article History
Received: 24.10.2026
Revised: 29.10.2026
Accepted: 03.11.2026

This article is published under the terms of the Creative Commons Attribution License 4.0.

INTRODUCTION

Plant breeding has entered a new era in which genomic information is becoming increasingly important for improving crop productivity, quality and adaptability. The development of DNA sequencing technologies has enabled researchers to understand the genetic composition of crop species and identify genes associated with important agricultural characteristics. Traditionally, these investigations have relied on a single reference genome, which serves as a standard sequence for identifying genetic variations and understanding the organization of plant chromosomes.

This limitation creates a problem known as reference bias, in which genetic analyses become more effective at detecting variations similar to the reference genome while overlooking sequences that are absent or highly divergent from it. Consequently, potentially useful genetic information may remain undiscovered. Pangenomics provides a solution by integrating genomic information from multiple individuals of the same species. Instead of representing a crop through one genome, it captures the collective genetic content and diversity of a population. Bayer *et al*. (2020) emphasized that plant pangenomes provide a broader representation of genetic diversity and can support breeding by identifying conserved and variable genes.

The concept has further expanded into super pangenomics, which incorporates genomic information from multiple related species within a genus or broader taxonomic group. This approach is particularly valuable because crop wild relatives often contain genetic resources that have been lost or reduced during domestication and modern breeding. These resources may provide genes for disease resistance, drought tolerance, salinity tolerance and other important agronomic characteristics (Khan *et al*., 2020).

Understanding the Concept of a Reference Genome

A reference genome is a representative DNA sequence used as a standard framework for studying the genome of an organism. It provides information about chromosome organization, gene locations, regulatory regions and other genomic features.

In plant breeding, reference genomes are used for several important purposes:

1. Identification of genes associated with economically important traits.
2. Detection of single nucleotide polymorphisms and other genetic variations.
3. Development of molecular markers for marker-assisted selection.
4. Identification of quantitative trait loci associated with yield and stress tolerance.
5. Understanding gene expression, evolution and domestication.
6. Supporting genome-wide association studies and genomic selection.

What Is a Pangenome?

A pangenome is the complete collection of nonredundant genetic sequences found across multiple individuals belonging to the same species. It provides a broader understanding of the genetic composition of a species by combining information from different varieties, accessions, landraces and breeding populations.

Core genome

The core genome consists of genes that are present in all or nearly all individuals included in a particular pangenome analysis. These genes generally perform essential biological functions involved in growth, development, metabolism and basic cellular activities. Core genes are important for understanding conserved biological processes and identifying genetic regions that remain relatively stable across different varieties.

Dispensable genome

The dispensable genome includes genes that are present in some individuals but absent in others. These genes may contribute to differences in disease resistance, environmental adaptation, plant architecture and other agronomic characteristics.

Private or accession-specific genes

Private genes are genes identified in particular individuals or accessions within the analysed population. They may represent unique genetic resources associated with specific geographical regions, evolutionary histories or environmental adaptations. These genes can be valuable for discovering novel traits and expanding the genetic diversity available to breeders. The identification of core and variable genomic components provides a more comprehensive understanding of genetic diversity and the evolutionary processes influencing crop improvement (Bayer *et al*., 2020).

From Pangenome to Super Pangenome

A conventional pangenome generally focuses on genetic diversity within a single species. However, crop improvement often requires genetic resources beyond the cultivated gene pool. Crop wild relatives and related species contain valuable genetic variations that may have disappeared or become less frequent during domestication and modern selection. Super pangenomics expands the pangenome concept by incorporating genomic information from multiple related species, particularly within the same genus. For example, a super pangenome of rice can potentially integrate genetic information from cultivated rice and related wild *Oryza* species. Similarly, a super pangenome of wheat may include cultivated wheat and related wild and ancestral species

Table 1: Major differences between reference genome, pangenome and super pangenome

Feature	Reference genome	Pangenome	Super pangenome
Genetic coverage	One representative individual	Multiple individuals of one species	Multiple related species and their accessions
Main objective	Provide a standard genomic sequence	Capture genetic diversity within a species	Capture broader genetic diversity across related species
Genetic information	Mainly reference sequences	Core and variable genes	Genus-conserved, variable and species-specific genes
Genetic diversity	Limited representation	Broad population-level diversity	Wider evolutionary and species-level diversity
Wild relatives	Generally not included	May be included if belonging to the target species	Specifically incorporates related wild species
Breeding application	Marker development and gene identification	Identification of useful alleles within a crop species	Discovery of novel genes from wild relatives for crop improvement
Major advantage	Provides a standard genomic framework	Reduces reference bias and improves gene discovery	Expands the available genetic resources for future breeding

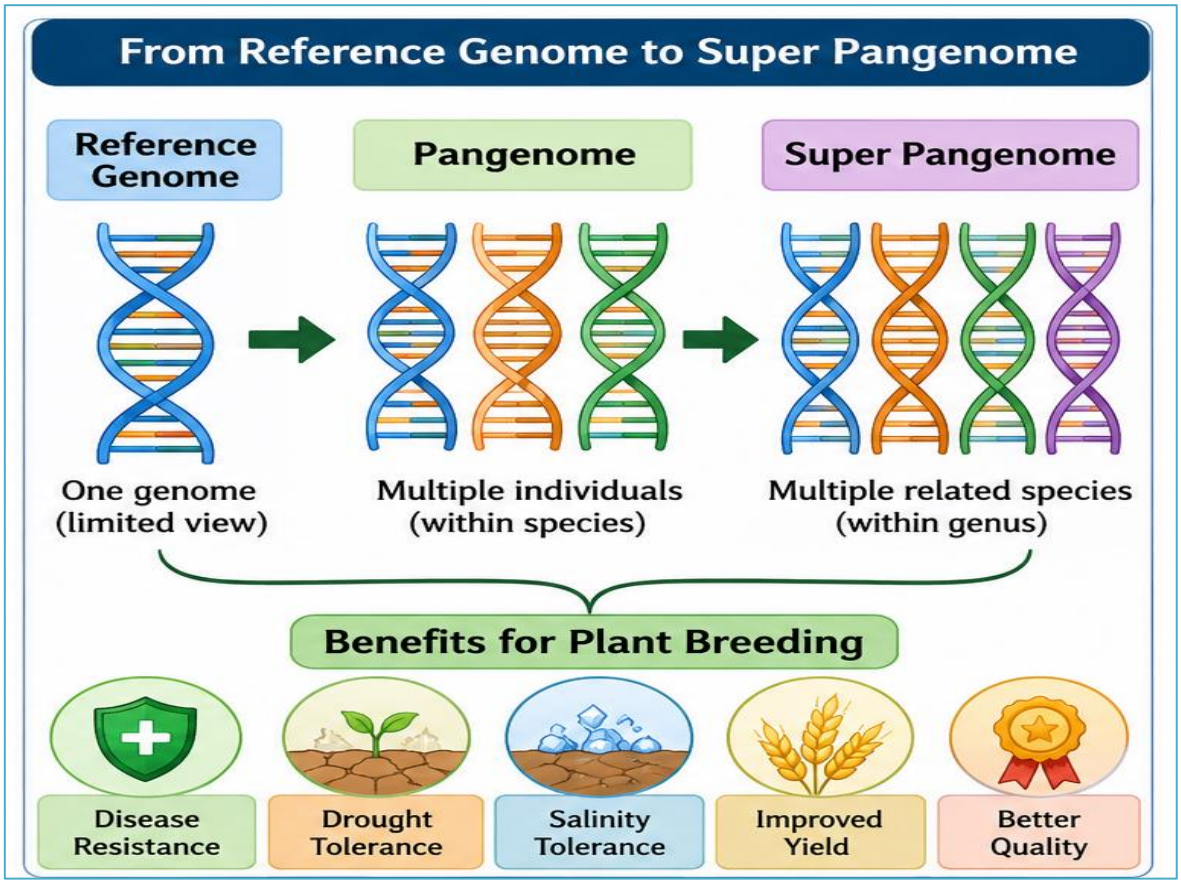


Fig. 2: Evolution from a single reference genome to a super pangenome for enhanced crop improvement.

Major Applications of Pangenomes and Super Pangenomes in Plant Breeding

Pangenomics has considerable potential for improving the identification and utilization of genetic resources in modern crop improvement programmes.

Discovery of novel genes and genetic variations

One of the most important applications of pangenomics is the identification of genes that are absent from conventional reference genomes. These novel genes may contribute to important characteristics such as disease resistance, stress tolerance, nutritional quality and yield improvement. A notable example is the tomato pangenome developed by Gao *et al*. (2019). By analysing 725 genetically and geographically representative tomato accessions, researchers identified 4,873 genes that were absent from the existing reference genome. The study also identified a rare allele of the TomLoxC gene promoter associated with fruit flavour.

Disease resistance breeding

Crop diseases caused by fungi, bacteria, viruses and other pathogens are major constraints on agricultural production. Pangenomes can help identify genes and genomic variations associated with resistance against important plant diseases. Variable genes may contain resistance-related sequences that are present in certain varieties or wild relatives but absent from other cultivated materials.

Development of climate-resilient crop varieties

Climate change is increasing the importance of developing crops that can tolerate drought, heat, salinity, waterlogging and other environmental stresses. Pangenomics can help identify genetic variations associated with adaptation to different environmental conditions. Wild relatives and traditional landraces often contain useful alleles associated with survival under adverse environments. By including these genetic resources in super pangenomes, breeders can identify candidate genes and genomic regions for improving stress tolerance.

Improvement of nutritional and grain quality

Pangenomics can contribute to the identification of genetic factors controlling nutritional composition, grain characteristics and quality-related traits. Important breeding objectives include improving protein content, mineral concentration, starch composition, oil quality and nutritional compounds.

Understanding crop domestication and evolution

Crop domestication has involved the selection of plants with desirable characteristics such as larger seeds, improved grain quality, reduced seed shattering, uniform maturity and higher productivity. During domestication and modern breeding, certain genetic variations may have been lost or reduced in frequency.

Improving hybrid breeding and genomic selection

Hybrid breeding requires the identification of parental lines that produce superior offspring. Pangenomes can improve parental selection by providing a broader understanding of genetic diversity and genomic relationships. Bayer *et al*. (2021) discussed the potential of combining pangenomics and machine learning with genomic selection in plant breeding. Such integration provides opportunities to improve predictive models by incorporating additional genomic variations rather than relying exclusively on markers identified using a single reference genome.

Importance of Super Pangenomes in Utilizing Crop Wild Relatives

Crop wild relatives are genetically related wild plant species that have not undergone the same degree of domestication and artificial selection as cultivated crops. They represent valuable sources of genetic diversity for modern plant

breeding. Many wild relatives possess characteristics that are less common in cultivated varieties, including resistance to insect pests, tolerance to extreme temperatures, adaptation to poor soils and resistance to emerging diseases.

Its major contributions include:

1. Identification of novel resistance genes from wild species.
2. Discovery of genetic variations associated with drought and heat tolerance.
3. Identification of ancestral genes lost during domestication.
4. Identification of useful alleles for improving nutritional quality.
5. Expansion of the genetic diversity available for hybridization and pre-breeding programmes.

Challenges and Limitations of Pangenomics and Super Pangenomics

Despite their considerable potential, pangenome and super pangenome technologies face several scientific and technical challenges.

- **High cost of genome sequencing:** Constructing high-quality pangenomes requires sequencing multiple genetically diverse accessions. Long-read sequencing, genome assembly and annotation can increase the overall cost of research.
- **Complex plant genomes:** Many crop species possess large genomes, repetitive DNA sequences and polyploid chromosome structures. These characteristics make genome assembly and comparison difficult. For example, wheat contains a complex hexaploid genome, which creates additional challenges in distinguishing closely related chromosome sequences and identifying genetic variations.

Bioinformatics and data storage requirements

Pangenome construction generates large volumes of genomic information that require advanced computational infrastructure. Researchers need suitable software, data storage systems, standardized pipelines and technical expertise to process and interpret these datasets.

Accurate functional annotation

Identifying a variable gene does not automatically establish its biological function. Some genes may have unknown functions or may be incorrectly annotated. Therefore, experimental validation, gene expression studies, functional genomics and controlled biological investigations are necessary to establish the contribution of candidate genes to important agronomic traits.

Representation of genetic diversity

An effective pangenome depends on the appropriate selection of accessions. If only a limited number of varieties are included, rare alleles and geographically restricted genetic variations may remain undiscovered. Similarly, super pangenomes require careful selection of related species and representative populations to ensure adequate coverage of genetic diversity.

Future Prospects of Pangenomics in Agriculture

- Development of crop-specific super pangenomes: More comprehensive genomic resources are needed for important crops such as rice, wheat, maize, soybean, pulses and horticultural crops.
- Integration of multi-omics information: Combining genomic information with transcriptomics, proteomics, metabolomics and phenomics can improve the understanding of gene functions and complex agricultural traits.
- Artificial intelligence-assisted gene discovery: Machine learning and deep learning can help identify candidate genes, structural variations and useful alleles from large pangenomic datasets.
- Precision breeding: Pangenomic information can support more accurate

selection of parental lines and the identification of genetic combinations suitable for specific breeding objectives.

- ❖ Conservation of genetic resources: Pangenomes can document valuable genetic diversity present in landraces and wild relatives, supporting their conservation and utilization.
- ❖ Climate-resilient agriculture: Super pangenomes can help identify genetic resources for developing crops with improved tolerance to drought, heat, salinity and emerging pests and diseases.

CONCLUSION

Pangenomes and super pangenomes represent an important advancement in modern plant genomics by overcoming the limitations of relying on a single reference genome. They provide broader access to genetic diversity, novel genes and useful variations present in cultivated varieties, landraces and wild relatives. Their integration with genomic selection, GWAS, artificial intelligence and genome editing offers new opportunities for developing high-yielding, disease-resistant and climate-resilient crop varieties. Ultimately, moving from a single reference genome towards comprehensive pangenomic resources will help breeders discover hidden genetic potential and accelerate sustainable crop improvement.

REFERENCES

Bayer, P. E., Golicz, A. A., Scheben, A., Batley, J. & Edwards, D. (2020). Plant pan genomes are the new reference. Nature Plants, 6(8), 914–920

Bayer, P. E., Petereit, J., Danilevicz, M. F., Anderson, R., Batley, J. & Edwards, D. (2021). The application of pangenomics and machine learning in genomic selection in plants. The Plant Genome, 14(3), e20112.

Gao, L., Gonda, I., Sun, H., Ma, Q., Bao, K., Tieman, D. M., Burzynski-Chang, E. A., Fish, T. L., Stromberg, K. A., Sacks, G. L., Thannhauser, T. W., Foolad, M. R., Diez, M. J., Blanca, J., Canizares, J., Xu, Y., van der Knaap, E., Huang, S., Klee, H. J., Fei, Z. (2019). The tomato pan genome uncovers new genes and a rare allele regulating fruit flavor. Nature Genetics, 51(6), 1044–1051.

He, W., Li, X., Qian, Q. & Shang, L. (2025). The developments and prospects of plant super pangenomes: Demands, approaches and applications. Plant Communications, 6(2), Article 101230.

Jayakodi, M., Shim, H. & Mascher, M. (2025). What are we learning from plant pangenomes? Annual Review of Plant Biology, 76, 663–686.

Khan, A. W., Garg, V., Roorkiwal, M., Golicz, A. A., Edwards, D. & Varshney, R. K. (2020). Super pangenome by integrating the wild side of a species for accelerated crop improvement. Trends in Plant Science, 25(2), 148–158.

Schreiber, M., Jayakodi, M., Stein, N. & Mascher, M. (2024). Plant pangenomes for crop improvement, biodiversity and evolution. Nature Reviews Genetics, 25, 563–577.