



Pangenomics is Rewriting the Reference Genome: How Multiple Crop Genomes Are Revealing Hidden Diversity for The Next Generation of Breeding

*Mohammed Shahid¹, Charith R.², Rahul G.K.³ and Meghana R. Sagar⁴

¹M.Sc. Scholar, Department of Plant Biotechnology, University of Agricultural Sciences, Bangalore, Karnataka, India

²Research Associate, Mahyco Private Limited, India

³PhD Scholar, Department of Plant Biotechnology, University of Agricultural Sciences, Bangalore, Karnataka, India

⁴PhD Scholar, ICAR - Indian Agricultural Research Institute, New Delhi, India

*Corresponding Author's email: mohammedzeshan26@gmail.com

The Problem: One Genome Cannot Represent a Crop

For decades, crop genomics has revolved around a deceptively simple idea: choose one high-quality genome and use it as the reference for studying an entire species. This approach transformed plant genetics, enabling researchers to map genes, discover SNPs, annotate genomes and accelerate marker-assisted breeding. But a single reference genome can never capture all the genetic diversity present across thousands of varieties, landraces and wild relatives. This limitation is becoming increasingly important as modern sequencing technologies reveal variation far beyond single-nucleotide changes. Insertions, deletions, duplications, inversions, copy-number changes and entire genes that are present in one variety but absent in another can all contribute to important agricultural traits. Pangenomics changes the question. Instead of asking how every crop differs from one reference genome, researchers can now examine the genomic landscape of an entire species.

From One Reference to Many Genomes

A pangenome is a collection of genome sequences representing multiple individuals within a species or, in some cases, a broader taxonomic group. Rather than representing a crop as one genome □ one reference, pangenomics represents it as many varieties + landraces + breeding lines + wild relatives □ one genomic diversity resource.

Beyond SNPs: The Hidden World of Structural Variation

Traditional crop genomics has often focused heavily on SNPs and small indels. But two varieties can differ by much larger genomic changes: large insertions and deletions, gene duplications, copy-number variation, inversions, translocations, and presence-absence variation of genes. These are collectively important components of structural variation (SV).

A single reference genome can make some of these variants difficult to detect because sequences that are completely absent from the reference have nowhere to align. Pangenomes provide a framework for identifying these missing pieces of diversity. The importance of this became particularly clear in barley. A 2024 study constructed a pangenome using 76 chromosome-scale long-read assemblies together with short-read data from 1,315 additional genotypes. The researchers identified structurally complex regions associated with disease resistance, plant architecture, nutrient release and grain traits. In other words, some of the genetic differences that matter to agriculture are not simply “one nucleotide versus another.” They can involve entire genomic structures.

Crop Pangenomes at Work

Wheat: Recovering Diversity Lost from Modern Breeding

A 2024 study analyzed 827 Watkins wheat landraces together with modern wheat cultivars. The researchers identified approximately 262 million high-quality SNPs, along with millions of indels and tens of thousands of copy-number variants. Importantly, many variants found in the landraces were absent from modern wheat.

Structural Variation Meets Wheat Breeding

Another recent wheat study examined 17 chromosome-level wheat cultivar assemblies and identified 249,976 structural variations, with almost half exceeding 5 kb. The researchers used these variations to investigate patterns associated with wheat habitat and breeding history.

Rice: When Structural Variants Influence Agronomic Traits

A graph-based rice super-pangenome constructed from 251 accessions, including cultivated and wild rice, revealed extensive structural and gene presence–absence variation. Researchers identified nearly 194,000 structural variants and demonstrated associations between particular variants, gene expression and agronomic traits.

Pangenomics for Climate-Resilient Crops

A 2025 maize study constructed a pangenome from 56 high-quality assembled genomes, including 25 germplasms selected for variation in drought resistance. The analysis revealed rare alleles and extensive regulatory diversity in drought-related genes and identified candidate genes associated with drought resistance.

The Rise of the Graph Genome

If several genomes contain different versions of the same genomic region, forcing all sequences onto one linear reference can introduce reference bias. A graph genome provides another solution. Instead of representing only one linear sequence, a graph can represent alternative paths corresponding to naturally occurring genomic alternatives.

From Pangenome to Breeding

The ultimate value of a crop pangenome is not the number of genomes assembled. It is the use of the diversity those genomes reveal.

Pangenomic resources can support:

- **Disease resistance:** Wild relatives and landraces may contain resistance-associated genes or structural variants absent from elite germplasm.
- **Drought and heat tolerance:** Rare alleles and regulatory variants can be investigated across germplasm adapted to different environments.
- **Yield improvement:** Structural variants affecting gene dosage or regulation can contribute to variation in yield components.
- **Nutritional quality:** Pangenomes can reveal genes and alleles associated with metabolic and nutritional traits.
- **Genomic selection:** A richer representation of genetic diversity can potentially improve the markers and haplotypes used to predict breeding values.
- **Genome editing:** Once useful variants are identified, pangenomic information can help breeders determine which genes or regulatory sequences are worth targeting.

The New Reference Genome

The concept of a reference genome is therefore changing. For years, the reference was essentially: “This is the genome we use to compare everything else against.” The emerging pangenomic view is closer to: “This is the genomic diversity available within the crop.” A breeder is not interested only in what exists in one reference variety. The valuable allele may be hiding in an old landrace, a wild relative or a geographically isolated cultivar. Pangenomics provides the computational framework for finding it.

The Horizon

The next phase of crop genomics is likely to combine long-read sequencing, chromosome-scale assemblies, structural-variant discovery, graph genomes, phenotyping, AI and genomic prediction. The direction is moving from: Reference genome □ Multiple genomes □ Pangenome □ Graph-based genomic diversity □ Trait-associated variation □ Breeding deployment

The central challenge will no longer simply be generating more genomes. It will be determining which parts of the enormous diversity captured by pangenomes are biologically meaningful, agronomically useful and practically deployable in breeding programmes.

As crop pangenomes continue to expand, the genome of a single elite cultivar may increasingly be viewed not as the reference, but as one point within a much larger landscape of agricultural diversity.

The future of crop genomics may therefore not be about finding a better single reference. It may be about building a reference that represents the diversity of the crop itself.

References

1. Schreiber, M., Jayakodi, M., Stein, N. et al. (2024). Plant pangenomes for crop improvement, biodiversity and evolution. *Nature Reviews Genetics*, 25, 563–577.
2. Jayakodi, M., Lu, Q., Pidon, H. et al. (2024). Structural variation in the pangenome of wild and domesticated barley. *Nature*, 636, 654–662.
3. Cheng, S., Feng, C., Wingen, L.U. et al. (2024). Harnessing landrace diversity empowers wheat breeding. *Nature*, 632, 823–831.
4. Jiao, C., Xie, X., Hao, C. et al. (2025). Pan-genome bridges wheat structural variations with habitat and breeding. *Nature*, 637, 384–393.
5. Yang, S., Wang, Y., Huang, Q. et al. (2025). A pangenome of maize provides genetic insights into drought resistance. *Nature Genetics*, 57, 2831–2841.