AGRI MAGAZINE
(International E-Magazine for Agricultural Articles)
Volume: 02, Issue: 12 (December, 2025)
Available online at http://www.agrimagazine.in
©Agri Magazine, ISSN: 3048-8656

Pangenomes: The Full DNA Library Transforming Crop Improvement

*A. Harshika Sri[1], P. R. Patel[2] and Vaishali Dangar[3]
[1]Ph.D. Research Scholar, Department of Genetics and Plant Breeding, B. A. College of Agriculture, AAU, Anand-388001 (Gujarat), India
[2]Ph.D. Research Scholar, Department of Genetics and Plant Breeding, N. M. College of Agriculture, NAU, Navsari-396450 (Gujarat), India
[3]M.Sc. Researcher, Department of Genetics and Plant Breeding, N. M. College of Agriculture, NAU, Navsari-396450 (Gujarat), India
*Corresponding Author's email: aharshikasri3009@gmail.com

Crop improvement has traditionally relied on a single reference genome to study genetic variation and guide breeding efforts. However, advances in genomics have revealed that no single genome can adequately represent the vast genetic diversity that exists within a crop species. Many important genes associated with stress tolerance, disease resistance, adaptation and quality traits are absent from standard reference genomes and remain hidden in landraces, traditional cultivars and wild relatives. The concept of the pangenome overcomes this limitation by integrating genomic information from multiple individuals to capture the complete gene repertoire of a species. By distinguishing between core genes shared by all individuals and variable genes present only in specific accessions, pangenomes provide deeper insights into structural variations such as gene presence–absence variation, duplications and large insertions or deletions that play a crucial role in trait expression. This expanded genomic perspective is reshaping plant breeding by enabling more precise gene discovery, marker development and targeted breeding strategies, including genomic selection and gene editing. Pangenomes therefore represent a transformative tool for harnessing hidden genetic diversity and developing climate-resilient, high-yielding and sustainable crop varieties for the future.

Introduction

For decades, plant genetics and breeding operated under the assumption that one "reference genome" could represent an entire crop species. That reference often derived from a single elite variety served as the foundation: researchers mapped traits, identified genes and bred for yield, disease resistance or quality based on that single genetic blueprint. But as scientists began comparing different varieties, landraces and wild relatives, a profound realisation emerged: one genome can never represent all the genetic potential hidden in a species. Different individuals, even of the same species can carry dramatically different genes. Some may possess unique variants that allow them to withstand drought, resist pests, or flourish under marginal soils. Others may bear genes that express desirable qualities like superior nutritional content or shelf-life. A lone reference genome inevitably misses many of these especially rare, region-specific or wild-derived genes. That's the root of the problem that led researchers to envision the pangenome: a "whole-species library" capturing all genes present across all individuals (or as many as feasible) rather than just one.

What exactly is a Pangenome and what it reveals?

The term "pangenome" (from Greek pan = all) was originally coined in microbial genetics to describe the full gene repertoire of all strains within a bacterial species (Tettelin *et al.*, 2005).

In plants, the idea was adopted because plant species often display high genetic and structural variation far beyond what a single genome can reveal (Zanini *et al.*, 2022).

A plant pangenome typically comprises two broad parts:

1. **Core genome**: genes shared by every individual sequenced these tend to code for basic, essential functions like growth, reproduction, metabolism, photosynthesis.
2. **Variable (or accessory) genome**: genes absent in one or more individuals these often contribute to traits of adaptation, stress response, disease resistance, environmental resilience, or novel traits like special flavors, colors, or metabolite profiles.

Because of this variation, pangenomes unmask structural differences not just the small DNA-letter changes (SNPs) but larger changes such as gene presence/absence, duplications, insertions/deletions, or inversions (Hurgobin & Edwards, 2017). These structural variants (SVs) often have major effects on traits. A pangenome is a comprehensive inventory the genetic "library" of a species that includes both common and rare, unique genes.

Why Pangenomes matter for crop improvement?

1. Capturing Hidden Genetic Diversity

When breeders rely on a single reference genome, they miss out on unique genes present in other varieties, landraces or wild relatives. For instance, wild or regional landrace populations might hold variants that enable drought tolerance, salinity resistance, or pest/disease resistance traits increasingly important under climate change or in marginal soils (Song et al., 2020). Pangenomes help reveal these hidden gems. By bringing together many genomes, scientists can map which individuals carry which genes enabling more informed selection and breeding. This helps especially for under-utilized or "orphan" crops that historically lacked genomic resources, expanding the possibility of improving crops vital for local food security but neglected in mainstream breeding.

2. Structural Variation: More than Just Letters

Most early genetic studies focused on single nucleotide differences (SNPs). While useful, SNPs often explain only a fraction of the variation in complex traits. Many important traits disease resistance, stress adaptation, yield potential can arise from structural changes: entire genes gained or lost, gene duplications, regulatory region changes, or large insertions/deletions.

Pangenomic studies make these structural variants visible. For example, pangenome studies in various crops have found that many "resistance genes" (R-genes) responsible for pathogen or stress defense lie in the variable genome. In some species, a large fraction of such R-genes were completely missing from the reference genome, but present in other accessions. Thus, pangenomes expand the palette of possible traits breeders can tap into far beyond what a single reference genome can suggest.

3. Enabling Precision Breeding, Genomic Selection & Gene Editing

With a rich catalog of genes and structural variants, breeders and geneticists can design more accurate molecular markers not only for SNPs but for presence/absence variants, copy-number variants, and structural changes to track traits in breeding populations.

Moreover, once beneficial genes are discovered whether from wild relatives, landraces or exotic varieties they can be more easily introduced into elite lines either through cross-breeding or modern methods like gene editing. This accelerates breeding and widens the pool of useful traits stress tolerance, nutrient quality, resilience.

4. Understanding Domestication, Evolution, and Biodiversity

Beyond immediate breeding applications, pangenomes provide insight into how domestication, human selection, and natural adaptation have shaped plant genomes over centuries. By comparing wild ancestors, landraces, and modern cultivars, researchers can see which genes were lost during domestication sometimes losing useful traits like stress tolerance or nutrient content in favor of yield. Pangenomes help trace these changes and offer a route to re-introduce lost traits.

Also, for polyploid species (those with multiple sets of chromosomes- like wheat), pangenomes help manage enormous genomic complexity and inter-individual variation, which single references cannot handle well.

Pangenomes Already Changing the Game

One of the most striking recent successes comes from rice. A large pangenome built from over 140 wild and cultivated rice accessions revealed nearly 3.87 billion base pairs of novel sequences not present in the classical reference genome (Zhao et al., 2018). This expanded "library" added tens of thousands of genes many linked to disease resistance and environmental adaptation.

Similarly, in crops such as wheat, tomato, soybean, canola, even orphan and under-utilized crops, pangenomes are being constructed giving breeders access to previously unknown genetic variation.

In tomato, for instance, pangenome and "super-pangenome" efforts (combining many cultivated and wild accessions) uncovered thousands of non-reference genes some potentially responsible for flavor, stress tolerance, or other desirable traits (Walkowiak et al., 2020; Gao et al., 2019).

Challenges and the Road Ahead

Despite their power, pangenomes are not a panacea. Constructing a pangenome requires sequencing many individuals, ideally representing the full diversity (wild relatives, landraces and cultivated varieties). For some minor or orphan crops, such diversity may not be well-documented or accessible.

The computational and bioinformatics challenges are also non-trivial. Combining many genomes, managing structural variation, building "graph-based" genomes rather than simple linear references these require advanced tools and expertise (Montenegro et al., 2017). Moreover, linking identified genes to actual agronomic traits remains a bottleneck: discovering a gene is one thing, but validating its effect under field conditions, ensuring it works across environments and genetic backgrounds, and integrating it into breeding pipelines takes time and effort.

Nevertheless, the trajectory is clear: as sequencing becomes cheaper, more genomes from diverse accessions are added and bioinformatics tools mature, pangenomes will become ever more central to plant breeding. The concept is now evolving further into the idea of a "super-pangenome" an inclusive genome resource for an entire genus rather than just a species, capturing even broader genetic variation.

Conclusion: A New Era of Data-Driven, Diversity-Driven Breeding

The pangenome paradigm is transforming how we view crop genetics. Rather than relying on a single "standard" genome, scientists now recognize that diversity within a species is its greatest strength. By capturing the full genetic repertoire all core, accessory and unique genes pangenomes provide breeders and geneticists with a vast toolkit for improvement: adaptation, resilience, nutrition, yield, and sustainability. As more crop species get pangenomes including both major staples and under-utilized local crops the potential grows more resilient agriculture, better crops for future climates and a reconnection of modern breeding with the rich genetic heritage of landraces and wild plants.

References

1. Bayer, P. E., *et al*. (2020). Plant pan-genomes are the new reference. Nature Plants, 6, 914–920.
2. Gao, L., *et al*. (2019). Tomato pan-genome reveals flavor genes. Nature Genetics, 51, 1044–1051.
1. 3. Golicz, A. A., *et al*. (2016). Towards plant pangenomics. Plant Biotechnology Journal, 14(4), 1099– 1105.
3. Hurgobin,B., & Edwards, D. (2017). SNP discovery using a pangenome. Biology, 6(2), 21.

4. Montenegro, J. D., *et al.* (2017). The pangenome of hexaploid wheat. The Plant Journal, 90, 1007–1013.
5. Song, J. M., *et al.* (2020). Pan-genome architecture of Brassica napus. Nature Plants, 6, 34–45.
6. Tettelin, H., *et al.* (2005). Genome analysis of multiple isolates reveals the microbial pan-genome. PNAS, 102(39), 13950–13955.
7. Walkowiak, S., *et al.* (2020). Multiple wheat genomes reveal global variation. Science, 368, 637–641.
8. Zanini, S. F., *et al*. (2022). Pangenomics in crop improvement. The Plant Genome, 15(1), e20177.
9. Zhao, Q., *et al*. (2018). Pan-genome analysis of rice. Nature Genetics, 50, 278–284.