

Pan-Genomics in Fruit Crops: Current Progress, Challenges and Future Directions

Anupama Roy¹, Sarika² and Mir Asif Iquebal²

¹*The Graduate School, ICAR-Indian Agricultural Research Institute, New Delhi*

²*ICAR-Indian Agricultural Statistics Research Institute, New Delhi*

Received 17 February 2026; Revised 21 April 2026; Accepted 21 April 2026

SUMMARY

Pan-genomics captures the full spectrum of genetic diversity within plant species and overcomes the limitations of single reference genomes. By integrating core, dispensable, and unique genomic components, pan-genomic approaches enable comprehensive detection of structural variants, presence–absence variations, and lineage-specific genes that underpin phenotypic diversity, domestication, and environmental adaptation. In plant and horticultural crops, particularly tropical and subtropical fruit species characterized by high heterozygosity, polyploidy, and clonal propagation, pan-genomics has revealed extensive hidden variation influencing fruit quality, disease resistance, and stress tolerance that was previously inaccessible using traditional genomics.

This review synthesizes the concepts, methodological advances, computational tools, and recent progress in plant pan-genomics, with a focused emphasis on tropical and subtropical fruit crops. We discuss pan-genome construction strategies, graph-based representations, and emerging resources, alongside key applications in trait discovery, evolutionary analysis, and molecular breeding. Current challenges, including limited accession diversity, incomplete functional annotation, and computational complexity, are critically examined. Finally, we highlight future directions toward population-scale, graph-based, and multi-omics–integrated pan-genomes, underscoring their potential to accelerate precision breeding and support the development of climate-resilient, high-quality fruit crop varieties.

Keywords: Pan-genomics; Tropical and subtropical fruit crops; Structural variation; Presence–absence variation (PAV); Graph-based genomes; Crop wild relatives; Molecular breeding; Genomic diversity; Long-read sequencing; Climate-resilient horticulture.

1. INTRODUCTION

Fruit crop genomics plays a vital role in enhancing agricultural productivity and sustainability by enabling precise genetic improvements tailored to fruits' unique biology. It addresses challenges like climate variability, pests, and consumer demands for better quality. Genomics analysis reduces breeding timelines from decades to years and enhances market competitiveness by improving shelf life, flavour, and stress tolerance, addressing post-harvest losses that cost billions annually (Roy *et al.*, 2025). Fruit crops are essential sources of vitamins, minerals, dietary fibre, and bioactive compounds, contributing significantly to human nutrition and health. Genomic studies have

illuminated the pathways responsible for synthesizing these nutrients, allowing targeted breeding for nutrient-dense and functional fruits. Integrating pan-genomics into fruit crop research has revolutionized the understanding of genetic diversity and adaptation.

Pan-genomics represents an advanced genomic framework that captures the complete genetic diversity across all strains or individuals within a species, moving beyond the constraints of a single reference genome to encompass both shared and unique genetic elements. This approach has revolutionized fields like microbiology, plant breeding, and human genomics by enabling more accurate variant detection, evolutionary studies, and personalized medicine applications.

Corresponding author: Sarika

E-mail address: aijaiswal@gmail.com

Single linear reference genomes fail to represent species-wide genetic diversity, causing reference bias and overlooking many structural variants and adaptive regions. As sequencing improved, it became clear that key loci linked to adaptation and disease resistance were missing from references, driving a shift toward pan-genome approaches that integrate multiple genomes to capture comprehensive intraspecies variation (Dutilh, 2016).

The aim of this review is to summarize the concept, methodological approaches, and recent advances in pan-genomics, with particular emphasis on its application in plant and horticultural crops. The review further highlights the status of pan-genomic studies in tropical and sub-tropical fruit crops, along with associated challenges and prospects for crop improvement.

2. CONCEPT AND EVOLUTION OF PAN-GENOMICS

The concept of the pan-genome originated in bacterial genomics, first formalized by Tettelin *et al.* (2005) in their comparative study of *Streptococcus agalactiae*. The pan-genome includes all genes present within all strains of a species, representing both shared and unique components. Since its introduction, the concept has expanded to plants, fungi, and even animal genomes as sequencing capacity increased (Tettelin *et al.*, 2005).

In pan-genomics, the core genome comprises genes present in all individuals or strains of a species, forming the essential functional backbone conserved across the population, typically encoding housekeeping functions like replication and metabolism. The dispensable (or accessory) genome includes genes absent in some strains but present in others, contributing to ecological adaptability; unique genes are strain-specific elements, often acquired via mobile elements, representing the most variable fraction. These categories collectively define the pan-genome as the union of all genes, with the core stabilizing species identity while the dispensable drives diversity and evolution (Matthews *et al.*, 2024; Oshkin *et al.*, 2020).

Beyond the core–dispensable framework, pan-genomes are often classified into soft-core genes (present in most accessions), shell genes (moderately frequent and associated with niche adaptation), and

cloud genes (rare, accession-specific, and often acquired through horizontal transfer). This hierarchical structure, commonly represented in pan-genome graphs, captures gradual variation across populations and has been shown to explain ecotype diversification, such as in *Prochlorococcus*, thereby improving genome evolution modeling and supporting applications in precision agriculture and disease surveillance (Matthews *et al.*, 2024; Sarawad *et al.*, 2025).

3. STRUCTURAL VARIATION AND GENOME PLASTICITY IN PLANTS

Structural variations (SVs), including presence–absence variations (PAVs), copy number variations (CNVs), inversions (segment reversals), and translocations (chromosomal relocations), which collectively contribute to genetic and phenotypic diversity (Yuan *et al.*, 2021; X. Zhang *et al.*, 2018). These SVs have been extensively studied in crop species such as rice, maize, wheat, and *Brassica napus*, where they play a pivotal role in adaptation, domestication, and breeding outcomes (Afsharyan *et al.*, 2025; Springer *et al.*, 2009; Y. Zhang *et al.*, 2024).

Transposable elements (TEs) are mobile DNA sequences capable of self-replication and insertion can create new insertions, deletions, or gene disruptions, contributing to genomic rearrangements and regulatory innovations (Hassan *et al.*, 2024). They facilitate the birth of novel genes and can alter gene expression under stress conditions, thus promoting adaptability. Similarly, whole-genome duplications (polyploidy) generate redundancy, providing raw material for evolutionary diversification through gene neofunctionalization and sub functionalization. In *Brassica napus* and *Bassia scoparia*, TE activity and segmental duplications have been linked to agronomically relevant traits such as stress tolerance and herbicide resistance, demonstrating how mobile elements and duplications underpin plant evolution and adaptation (Afsharyan *et al.*, 2025; Hall *et al.*, 2025).

After whole-genome duplication, plant genomes commonly undergo fractionation, marked by biased gene loss and reorganization across duplicated regions. Redundant gene copies are progressively removed or diversified, leading to sub-genome asymmetry and functional divergence while retaining essential genes. In polyploid crops such as *Brassica napus* and wheat,

fractionation is often biased toward one parental sub-genome and is accompanied by rapid genomic restructuring and epigenetic changes. This process enhances genome plasticity and facilitates the evolution of complex traits, enabling plants to adapt to diverse and changing environments (Afsharyan *et al.*, 2025; De Oliveira *et al.*, 2020; Weiss-Schneeweiss *et al.*, 2013).

Structural variants are increasingly recognized as major determinants of agronomic traits in plants, influencing yield, stress response, flowering time, and disease resistance. PAVs and CNVs often affect key regulatory genes or biosynthetic pathways linked to these traits. For instance, CNVs in flowering-time genes like *BnFLC* in *Brassica napus* and *SBE1* in sorghum have been linked to phenological diversity and adaptation to distinct environments (Afsharyan *et al.*, 2025). Large-scale SV analyses in rice and maize demonstrate that SV-enriched genomic regions frequently harbour genes responsible for stress responses and developmental regulation, underscoring the functional importance of SVs in crop improvement programs (Fuentes *et al.*, 2019; Springer *et al.*, 2009).

4. TRANSITION FROM LINEAR GENOMICS TO PAN-GENOMICS

Reference bias arises when reads from diverse genomes are mapped to a single linear reference, leading to misalignment of non-reference alleles, under-detection of structural variants like insertions/deletions (up to 20% missed in crops), and skewed variant frequencies that distort population genetics and GWAS results. In microbial and plant studies, this bias systematically excludes accessory genes and haplotype-specific paths, inflating false negatives in regions of high divergence, as demonstrated in comparative analyses of bacterial isolates where pan-genome graphs recovered 15-30% more variants (Matthews *et al.*, 2024; Urhan & Abeel, 2021).

Capturing the full genetic diversity of a species is essential for understanding evolution, adaptation, and phenotype variation. Since a single genome cannot represent all allelic, structural, and PAVs across a population, the pan-genome concept was introduced to integrate all genetic sequences present within a species (Loegler *et al.*, 2025).

Pan-genome capture a comprehensive spectrum of genetic variation, including Single Nucleotide Polymorphisms (SNPs), CNVs, inversions, and PAVs, enabling more accurate association studies and functional annotations. In plants, variant detection has revealed hidden variations influencing agronomic traits, stress responses, and adaptation to diverse environments. Furthermore, pan-genomes facilitate the identification of lineage-specific genes and evolutionary events, providing a foundation for studying genome dynamics at population and species levels (Du *et al.*, 2025; Munir *et al.*, 2020). Ultimately, pan-genomics enhances the resolution and accuracy of genomic analyses by bridging population-level variation with functional genomics (Loegler *et al.*, 2025; Naithani *et al.*, 2023).

By identifying the complete repertoire of genes across cultivars and wild relatives, breeders can access previously unexploited genetic variation for marker-assisted selection and genomic prediction. Moreover, pan-genome approaches facilitate the development of super-pangenomes, integrating hundreds of *de novo* assemblies to capture interspecies variation within crop genera. These advances help breeders design crops with enhanced adaptability and productivity under changing climates (Du *et al.*, 2025; Munir *et al.*, 2020; Tong *et al.*, 2025).

5. APPROACHES FOR PAN-GENOME CONSTRUCTION

Pan-genome construction employs diverse computational strategies to integrate multiple genomes, evolving from reference-dependent alignments to *de novo* and graph-based models that capture core and variable sequences across species. These approaches balance comprehensiveness, computational demands, and accuracy, with iterative and map-to-pan methods suiting short-read data while graphs excel in long-read resolutions. The iterative mapping and assembly approach is one of the earliest and most widely used methods for pan-genome construction. It involves aligning short reads from multiple individuals or accessions to a reference genome, identifying novel genomic regions that fail to align, and then assembling these unmapped reads to expand the reference genome iteratively (H. Hu *et al.*, 2024; Jiang *et al.*, 2024). This method progressively

incorporates genetic variation across individuals, capturing PAVs and novel genes absent in the reference. While it provides an efficient and computationally less demanding approach, it is highly dependent on the quality of the initial reference genome and may miss complex structural variants. Despite its limitations, iterative assembly remains valuable for species with relatively low diversity or limited genomic resources (Glick & Mayrose, 2023).

The map-to-pan approach enhances reference-based methods by incorporating novel sequences from unmapped reads into a composite pan-reference, enabling efficient detection of polymorphisms and PAVs while preserving reference coordinates. However, it remains prone to reference bias and may inadequately represent highly divergent or structurally complex genomic regions (H. Hu *et al.*, 2024; Jiang *et al.*, 2024). Thus, while computationally efficient, map-to-pan methods may be best suited for moderately variable species with high-quality references.

De novo assembly builds each genome independently without a reference, providing an unbiased approach to capture novel genes, structural variants, and repetitive regions. Comparing these assemblies identifies core and accessory sequences that define the pan-genome. While this method offers a comprehensive view of genomic diversity, it is computationally demanding and relies on high-coverage long-read sequencing for accurate assembly (Jiang *et al.*, 2024; Urhan & Abeel, 2021). Recent improvements in long-read sequencing technologies and assembly tools have significantly enhanced the feasibility of large-scale *de novo* pan-genome construction, allowing population-scale analysis of structural variation.

Graph-based pan-genomes integrate multiple genomes into a single graph rather than a linear reference, with nodes representing genomic segments and edges capturing alternative variants or structural rearrangements. This structure enables variation-aware read mapping and minimizes reference bias by allowing alignment across multiple genomic paths (Andreace *et al.*, 2023; Matthews *et al.*, 2024). Graph-based pan-genomes surpass linear references in genotyping accuracy and scalability, enabling bias-free read mapping, while long-read technologies improve repeat resolution. However, despite these advantages, graph-

based approaches remain computationally intensive and depend on advanced visualization and alignment tools that are still evolving (H. Hu *et al.*, 2024). A comparative overview of the different approaches for the pan-genome assembly methods is given in the Table 1.

Table 1. Comparison of Pan-genome Assembly Methods

Feature	Iterative mapping and assembly	Map-to-pan	<i>De novo</i> assembly	Graph pan-genome
Reference	Reference-based	Reference-based	Reference-free	Reference-free
Assembly level	Short reads	Contigs	Genomes	Genomes
Alignment approach	Iterative alignment	Genome-wide alignment	Genome-wide alignment	Genome-wide alignment
Sequencing depth requirements	Low depth	Low depth	High depth	High depth
Scope of application	Large-scale group	Large-scale group	Few individuals	Few individuals
Sequence storage	Linear	Linear	Linear	Graphic entity
Computing resource requirements	Low	Higher (in case of large genomes)	Higher	Higher
Analysis of SVs	Only SNPs and indels can be identified	Inter individual CNVs	Large SVs	Some highly divergent SVs
Cost	Low	Medium	Higher	Higher
Limitation	The sequence fragment is too short to analyze CNV	Not suitable for scale analysis of complex genomes	The genome needs to be annotated, which would lead to bias	Visualizing storage is difficult and ambiguity may arise during sequence alignment
Example	<i>Triticum aestivum</i> (Montenegro Cabrera, 2018); <i>Sorghum bicolor</i> (Ruperao <i>et al.</i> , 2021)	<i>Solanum lycopersicum</i> (Gao <i>et al.</i> , 2019); <i>Gossypium herbaceum</i> (J. Li <i>et al.</i> , 2021)	<i>Oryza sativa</i> (Schatz <i>et al.</i> , 2014); <i>Brassica napus</i> (Song <i>et al.</i> , 2020)	<i>Setaria</i> (He <i>et al.</i> , 2023); <i>Pennisetum glaucum</i> (Raza <i>et al.</i> , 2023)

6. COMPUTATIONAL TOOLS AND RESOURCES FOR PLANT PAN-GENOMICS

Plant pan-genomics depends on advanced computational tools for genome assembly, annotation, storage, and visualization to manage multiple complex

genomes, particularly those with high repeat content and polyploidy. The field is rapidly evolving, with graph-based approaches increasingly favoured for more comprehensive and unbiased genome representation.

Recent advances in long-read sequencing, hybrid assembly strategies, and telomere-to-telomere (T2T) assemblies have greatly improved the accuracy of plant pan-genomes, enabling near-complete and haplotype-resolved assemblies of complex crops. Alongside widely used alignment tools such as Minimap2, BWA-MEM, and NGMLR, specialized pan-genome tools support large-scale comparative analyses, allowing more precise detection of structural variants and gene presence–absence variations. An overall comparative pan-genome tools are provided in Table 2.

Table 2. Comparison of Pan-genome Assembly Tools

Tool Name	Read Type	Approach	Key Features
MaSuRCA (Zimin <i>et al.</i> , 2013)	Short/Long (hybrid)	Iterative mapping & assembly	Pools unmapped reads, metagenome-aware, low memory (~5GB), high precision for novel sequences
Hifiasm (Cheng <i>et al.</i> , 2021)	Long (PacBio HiFi, ONT)	<i>De novo</i> haplotype-resolved	Tri-HIFI mode for polyploids, phased assemblies, high contiguity (N50 >10Mb)
Flye (Kolmogorov <i>et al.</i> , 2019)	Long (ONT, PacBio)	<i>De novo</i>	Repeat-aware, polishing support, robust to low coverage
Canu (Koren <i>et al.</i> , 2017)	Long (PacBio, ONT)	<i>De novo</i>	Error-corrected overlaps, scalable for large genomes
EUPAN (Hu <i>et al.</i> , 2017)	Short/Long	Map-to-pan	Assembly-first then align to reference, high recall for insertions
PGGB (Garrison <i>et al.</i> , 2023)	Long (ONT, PacBio)	Graph-based	All-vs-all alignment, smooths noisy graphs, SV calling integration
Minigraph-Cactus (Hickey <i>et al.</i> , 2024)	Long	Graph-based	Reference-free graph from alignments, handles polyploidy
Panaroo (Tonkin-Hill <i>et al.</i> , 2020)	Short (Illumina)	<i>De novo</i> gene-based	Filters contaminants, gene family clustering

Accurate annotation of plant pan-genomes is vital to distinguish between core, accessory, and unique

genes, but is often hindered by reference bias and incomplete functional annotation pipelines. Traditional annotation pipeline such as, MAKER, Augustus depend on a single reference, leading to under representation of lineage-specific genes (Campbell *et al.*, 2014; Stanke *et al.*, 2006). Comparative approaches that integrate ab initio, homology-based, and RNA-seq-guided annotations have improved accuracy but still face challenges in polyploid plants with high gene redundancy. To address bias, pan-genome tools such as PanTools and PGAP (Pan-Genome Analysis Pipeline) incorporate multi-genome annotation and homology detection to capture all gene variants (Bu *et al.*, 2025; Jonkheer *et al.*, 2022). However, annotation consistency remains a major issue in cross-species pan-genomes, where ortholog detection can be confounded by gene family expansion or contraction, resulting in inaccurate clustering of homologous genes.

The massive data generated by pan-genomic analyses pose significant storage and visualization challenges. Traditional file formats (FASTA, VCF, GFF) are insufficient for multi-genome representations, leading to the development of more scalable formats like GFA (Graphical Fragment Assembly) and HDF5-based pan-genome databases (Jayakodi *et al.*, 2021; Naithani *et al.*, 2023). Visualization tools such as PanTools Viewer, and JBrowse2 enable interactive exploration of multi-genome alignments and annotations (Diesh *et al.*, 2023; Sheikhizadeh *et al.*, 2016). However, scalability and usability remain issues when visualizing large graph-based data, especially for polyploid crops.

7. PAN-GENOMIC STUDIES IN HORTICULTURAL AND FRUIT CROPS

Pan-genomic studies in horticultural and fruit crops have progressed from staple crops to diverse fruits, revealing extensive PAVs and SV that underpin traits like yield and stress tolerance. These efforts highlight both foundational insights and persistent challenges in capturing varietal diversity.

Early pan-genome studies established foundational frameworks in soybean, rice, and Brassica, demonstrating open gene pools where dispensable genes comprised 20-30% of the total. Soybean's PanSoy from 204 cultivated accessions identified millions of SVs and confirmed PAVs in trait genes via long-read validation, while rice pangenomes captured

subpopulation-specific contigs absent in Nipponbare, enhancing GWAS for grain traits (Torkamaneh *et al.*, 2021). Comparative chart for the up-to-date pangenome studies and their key findings in fruit crops is given in Table 3

Pan-genomics has recently been applied to tropical and subtropical fruit crops, revealing extensive intraspecific genetic diversity important for fruit quality, disease resistance, and environmental adaptation. Studies in crops such as mango, citrus, banana, pineapple, and tomato have uncovered high heterozygosity, transposable element-driven genome plasticity, and abundant SVs and PAVs influencing traits like flavour, fruit size, acidity, ripening, and hybrid vigour. Wild-domesticated comparisons often reveal 10–20% novel genes, providing valuable resources for marker development despite challenges posed by polyploidy. Pangenome construction over the year with the rapid sequencing technologies is shown in Fig. 1.

Overall, these fruit crop pan-genomes expose thousands of structural variants, novel genes, and allelic differences that shape key agronomic traits, including stress tolerance and disease resistance. Together, they demonstrate the power of pan-genomics to elucidate domestication processes and support molecular breeding for quality and resilience in fruit crops across diverse environments.

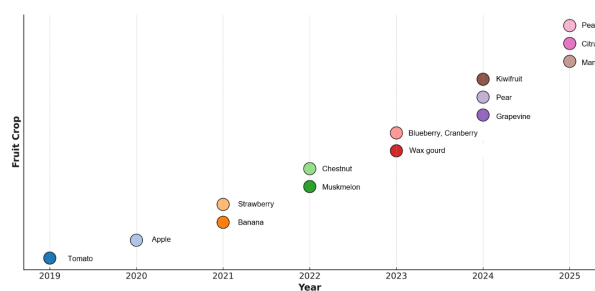


Fig. 1. Timeline showing the year-wise expansion of pan-genome studies across major fruit crops, reflecting increasing crop diversity and adoption of pan-genomic approaches over time

Pan-genome browsers dedicated exclusively to fruit crops remain scarce, with most studies relying on general plant portals or study-specific viewers rather than standalone fruit pan-genome interfaces. PlantPan provides the broadest multi-species fruit coverage, featuring gene-centric PAV matrices and ortholog clusters for few fruit crops (Jiang *et al.*, 2025). The Banana Genome Hub employs Panache for Musa-

Table 3. Summary of pangenome studies in fruit crops and their key genomic findings

Fruit	Findings	Reference
Tomato	Built tomato pan-genome (725 accessions, +4,800 novel genes). Found domestication-lost genes (especially disease resistance). Identified rare TomLoxC allele regulating fruit flavour. Rare allele mostly absent in modern cultivars but present in wild relatives.	(Gao <i>et al.</i> , 2019)
Apple	Built apple pan-genome with phased diploid assemblies. Discovered ~3,000 novel genes. Traced domestication history and allele-specific expression. Linked PAVs to trait-associated alleles.	(X. Sun <i>et al.</i> , 2020)
Banana	Constructed <i>Musa-Ensete</i> cross-genera pan-genome. Identified extensive PAVs and genomic variation. Revealed high divergence between the two genera.	(Rijzaani <i>et al.</i> , 2022)
Strawberry	Tracked evolutionary dynamics of gene families. Linked PAVs to key agronomic traits. Highlighted domestication-driven gene changes.	(Qiao <i>et al.</i> , 2021)
Muskmelon	Identified ~5,000 novel genes and abundant PAVs. Linked SVs/PAVs to fruit morphology traits. Provided genetic basis for trait-associated variation.	(Y. Sun <i>et al.</i> , 2022)
Chestnut	Built pan-genome of three major Chinese chestnut varieties. Identified gene family expansions and contractions. Detected extensive PAVs across varieties. Linked homologous genes to agronomic traits.	(G. Hu <i>et al.</i> , 2022)
Wax Gourd	Identified ~1,000 novel genes. Discovered resistance gene analogs (RGAs). Linked PAVs to fruit quality and trait variation.	(Yang <i>et al.</i> , 2023)
Blueberry, Cranberry	Constructed blueberry and cranberry pan-genomes. Identified extensive gene PAVs across accessions. Revealed candidate genes for fruit quality and stress tolerance.	(Yocca <i>et al.</i> , 2023)
Grapevine	Built grapevine pan-genome (hundreds of accessions). Identified >20,000 SVs and thousands of PAVs. Linked variants to disease resistance and fruit flavour traits. Provided toolkit for genomic breeding in <i>Vitis</i> .	(Z. Liu <i>et al.</i> , 2024)

Pear	Built haplotype-resolved T2T assemblies and pan-genome graph of pear. Identified extensive PAVs and SVs across accessions. Revealed diverse allele-specific expression patterns. Linked structural variants to fruit quality traits (flavour, texture).	(Q. Li <i>et al.</i> , 2024)
Kiwifruit	Built graph-based kiwifruit pan-genome. Identified SV underlying key traits. Found SVs regulating fruit degreening during ripening. Resource enables improved fruit quality breeding.	(Wang <i>et al.</i> , 2024)
Citrus	Constructed citrus pan-genome focusing on germin-like proteins (GLPs) gene family. Identified extensive gene family diversity and PAVs across species. GLP genes showed strong roles in disease resistance. Provided genomic insights for citrus improvement and stress resilience.	(Tahir ul Qamar <i>et al.</i> , 2025)
Mango	Revealed thousands of PAVs and SVs missed by single reference. Showed strong reference bias affecting population genomics. Provided resource for mango trait discovery and breeding.	(Ahmad <i>et al.</i> , 2025)
Peach	Built graph pan-genome of peach. Identified SVs and PAVs linked to malate content. Found NAC transcription factors regulating fruit acidity. Provided genetic basis for flavour and quality improvement.	(Chen <i>et al.</i> , 2025)

specific pan-gene heatmaps across accessions (Droc *et al.*, 2013). Citrus, kiwifruit, and tomato have their separate pangenome browsers (B. Li *et al.*, 2025; H. Liu *et al.*, 2022; Zhou *et al.*, 2022). While crop summaries embedded pan-modules in JBrowse for citrus and grapevine, underscoring that true fruit pan-genome browsers lag, often limited to offline ODGI tools or supplementary data (Guarracino *et al.*, 2022; Naithani *et al.*, 2023).

8. APPLICATIONS OF PAN-GENOMICS IN FRUIT CROP IMPROVEMENT

Fruit crop pan-genomics revolutionizes crop improvement by capturing full genetic diversity beyond single references, enabling discovery of trait-associated variants and accelerating precision breeding. This approach addresses genetic bottlenecks in elite varieties, integrating core/accessory genes for yield, quality, and resilience.

Pan-genomics has enabled the discovery of novel genes and allelic variants that were previously undetectable using single-reference genomes. By integrating genomes from multiple accessions, researchers can capture the full spectrum of genetic diversity, including PAVs and SVs that often underlie agronomic traits. For example, the *Setaria* (foxtail millet) pan-genome uncovered over 73,000 gene families and revealed key yield-related alleles such as *SiGW3*, whose promoter variation influences expression and productivity (He *et al.*, 2023)

Pan-genomics enables detailed analysis of evolution and domestication by revealing genomic changes underlying crop adaptation. For example, the spinach pan-genome uncovered more than 193,000 structural variants linked to key domestication traits such as sex determination, bolting time, and seed dormancy (She *et al.*, 2024). Crop wild relatives (CWRs) are rich sources of genetic diversity that can improve the resilience and productivity of modern cultivars. Pan-genomic approaches facilitate the identification and introgression of valuable alleles from wild species into cultivated crops. By enabling Pan-GWAS and graph-based analyses, pan-genomics is reshaping breeding strategies, allowing more accurate mapping of complex traits and supporting precision and predictive breeding.

Pan-genomics is increasingly moving from a discovery tool to a practical component of modern breeding by integrating structural variation data with genome-wide association studies (GWAS), genomic selection (GS), and emerging Artificial Intelligence (AI) -based prediction models. Unlike single-reference genomes, pan-genomes capture a broader spectrum of genetic diversity, including SVs and PAVs, which are often responsible for key agronomic traits but remain undetected in conventional analyses (Jayakodi *et al.*, 2021; Du *et al.*, 2025).

In breeding programs, pan-genome-enabled GWAS (Pan-GWAS) enhances trait mapping by incorporating non-reference sequences and structural variants. For example, the tomato pan-genome identified the *TomLoxC* allele associated with fruit flavour that was lost during domestication, providing a direct target for reintroduction into elite cultivars (Gao *et al.*, 2019). Similarly, graph-based pan-genomes have improved locus discovery and reduced missing heritability in

tomato and other crops (Zhou *et al.*, 2022), while in grapevine and citrus, pan-genomic analyses have revealed structural variants and gene family expansions linked to disease resistance and fruit quality traits, supporting their direct use in marker-assisted breeding (Liu *et al.*, 2024; Tahir ul Qamar *et al.*, 2025).

Pan-genomics also strengthens genomic selection by improving marker representation and predictive accuracy. Incorporating SVs and PAVs into GS models captures functional variation beyond SNPs, leading to better prediction of complex traits, particularly in crops with high structural diversity (Ruperao *et al.*, 2021; Jayakodi *et al.*, 2021). This is especially valuable for perennial fruit crops, where early and accurate selection can significantly reduce breeding cycles and improve genetic gain.

Furthermore, the integration of pan-genomic data with AI and machine learning approaches is enabling next-generation predictive breeding. By combining pan-genomic variants with multi-omics and phenotypic datasets, AI models can better capture complex genotype–phenotype relationships and environmental interactions (Tong *et al.*, 2025; Du *et al.*, 2025). Together, these advances support an integrated breeding framework in which pan-genomes inform variant discovery (Pan-GWAS), enhance prediction (GS), and enable data-driven selection (AI), ultimately accelerating the development of high-quality, climate-resilient fruit crop varieties.

9. CHALLENGES AND RESEARCH GAPS

Despite major advances in pan-genomics for staple cereals and legumes, fruit and horticultural crops remain underrepresented in pan-genomic research. This gap largely arises from the biological complexity of many fruit crops, including polyploidy, high heterozygosity, clonality, and large genome sizes, which complicate genome assembly and accurate variant detection. Perennial growth habits further limit the availability of diverse, well-characterized germplasm for large-scale sequencing. Consequently, many economically and nutritionally important fruit crops particularly those from tropical and subtropical regions lack comprehensive pan-genomic resources needed for trait discovery and modern breeding.

Horticultural crops, particularly perennial fruit trees, and clonally propagated species remain underrepresented in pan-genomics compared to annual grains, limiting insights into domestication sweeps and trait loci. Wild relatives (CWRs), critical for resilience alleles, are scarcely included; for instance, banana's seven-accession pan-genome overlooks *Musa balbisiana* diversity essential for *Fusarium* resistance (H. Hu *et al.*, 2025; Roy *et al.*, 2025; Sabety *et al.*, 2024). Expanding pan-genomic efforts to diverse fruit species is thus essential to represent the full spectrum of horticultural diversity.

A major limitation of current fruit crop pan-genomes is the narrow representation of cultivars, as many studies focus primarily on elite or commercial varieties while neglecting landraces, wild relatives, and local ecotypes that contain valuable adaptive and novel genetic variation. This restricted sampling leads to an underestimation of structural and presence–absence variations, resulting in a biased view of species-wide diversity. In addition, although whole-genome sequencing has advanced rapidly, limited integration of transcriptomic data has constrained functional annotation, with most fruit pan-genomes lacking extensive RNA-Seq resources to link genomic variation with gene expression, development, and stress responses.

User-friendly pan-genome resources for fruits are scarce, with data confined to supplements or project-specific JBrowse rather than centralized portals like PlantPan; visualization tools (Odgi, Panache) require expertise, limiting breeder access to SV/PAV queries (Guarracino *et al.*, 2022; Jiang *et al.*, 2025). No comprehensive fruit pan-genome hub exists hindering federated analysis across mango-citrus-grapevine; future needs include AI-phased graphs and phenotype-linked databases.

10. FUTURE PERSPECTIVES AND CONCLUSIONS

Pan-genomics has become an essential approach for comprehensively representing genetic diversity within plant species, addressing the shortcomings of relying on a single linear reference genome. By encompassing core, variable, and accession-specific genomic regions, pan-genomic frameworks offer a more complete view of genome organization, evolutionary dynamics, and

functional diversity. This is especially valuable for plant and horticultural crops that exhibit high levels of genetic variation and complex genome structures. In recent years, pan-genomic studies of tropical and subtropical fruit crops have uncovered extensive SVs and PAVs that were previously undetectable, shedding new light on the genetic basis of important traits, domestication processes, and adaptation to diverse environments.

Future progress in pan-genomics will depend on its integration with artificial intelligence (AI) to enable predictive breeding. A key priority is the development of AI-ready pan-genomic datasets that combine SVs and PAVs, and gene presence/absence matrices with high-quality phenotypic and environmental data. Graph-based pan-genomes provide an effective framework for this integration, allowing machine learning models to capture complex and non-linear genotype–phenotype relationships that are often missed in SNP-based analyses.

For practical implementation, breeders can adopt a stepwise pipeline: (i) construct diverse pan-genomes including wild relatives, (ii) identify trait-associated variants through Pan-GWAS, (iii) incorporate SVs and PAVs into genomic selection models, and (iv) apply AI or deep learning models to predict trait performance and select superior genotypes. Developing user-friendly platforms that integrate pan-genome data with AI-based prediction tools will be critical to enable routine use in breeding programs and accelerate the development of climate-resilient, high-quality fruit crops (Zhou *et al.*, 2022; Liu *et al.*, 2024).

Nevertheless, the application of pan-genomics in fruit crops faces notable challenges, including insufficient representation of diverse accessions, complications arising from heterozygosity and polyploidy, and the substantial computational demands of constructing and visualizing large pan-genomes. Overcoming these limitations will depend on incorporating a wider range of wild relatives and landraces, expanding the use of long-read sequencing technologies, and developing more efficient and accessible analytical platforms. Looking ahead, pan-genomic research is expected to evolve toward population-scale, graph-based models integrated with functional annotation and multi-omics data. These advancements will strengthen the role of

pan-genomics in trait discovery, molecular breeding, and the development of climate-resilient fruit crops, supporting sustainable horticultural improvement in the future.

ACKNOWLEDGEMENT

The authors also acknowledge the research support received under the research project "NNP (Agri-Genomic repository and Intelligent Analytical System) of ICAR Indian Agricultural Statistics Research Institute, New Delhi (Project No. BT/ PR40191/ BTIS/137/82/2023)" funded by the Department of Biotechnology (DBT), Government of India.

REFERENCES

- Afsharyan, N.P., Golicz, A.A., and Snowdon, R.J. (2025). Pangenomic structural variant patterns reflect evolutionary diversification in *Brassica napus*. *Genome Biology*, **26**(1), 381. <https://doi.org/10.1186/s13059-025-03833-x>
- Ahmad, B., Su, Y., Hao, Y., Razzaq, T., Arshad, R., Zhang, Y., Zhang, Y., Wang, X., Huang, G., Su, X., Hou, T., Li, C., Yang, X., Li, C., Chu, Z., Wang, Q., Zhang, Y., Jin, Z., Xu, Q., ... and Garrison Tian, X. (2025). Mango pangenome reveals dramatic impacts of reference bias on population genomic analyses. *Horticulture Research*, **12**(9), uhaf166. <https://doi.org/10.1093/hr/uhaf166>
- Andreace, F., Lechat, P., Dufresne, Y., and Chikhi, R. (2023). Comparing methods for constructing and representing human pangenome graphs. *Genome Biology*, **24**(1), 274. <https://doi.org/10.1186/s13059-023-03098-2>
- Bu, C., Zhang, H., Zhang, F., Liang, W., Gao, H., Zhao, J., Lv, F., Xue, R., Liu, Q., Zhang, Z., Jin, Z., and Xiao, J. (2025). PGAP2: A comprehensive toolkit for prokaryotic pan-genome analysis based on fine-grained feature networks. *Nature Communications*, **16**(1), 9865. <https://doi.org/10.1038/s41467-025-64846-5>
- Campbell, M.S., Law, M., Holt, C., Stein, J.C., Moghe, G.D., Hufnagel, D.E., Lei, J., Achawanantakun, R., Jiao, D., Lawrence, C.J., Ware, D., Shiu, S.-H., Childs, K.L., Sun, Y., Jiang, N., and Yandell, M. (2014). MAKER-P: A Tool Kit for the Rapid Creation, Management, and Quality Control of Plant Genome Annotations. *Plant Physiology*, **164**(2), 513-524. <https://doi.org/10.1104/pp.113.230144>
- Chen, W., Xie, Q., Fu, J., Li, S., Shi, Y., Lu, J., Zhang, Y., Zhao, Y., Ma, R., Li, B., Zhang, B., Grierson, D., Yu, M., Fei, Z., and Chen, K. (2025). Graph pangenome reveals the regulation of malate content in blood-fleshed peach by NAC transcription factors. *Genome Biology*, **26**(1), 7. <https://doi.org/10.1186/s13059-024-03470-w>
- Cheng, H., Concepcion, G. T., Feng, X., Zhang, H., and Li, H. (2021). Haplotype-resolved *de novo* assembly using phased assembly graphs with hifiasm. *Nature Methods*, **18**(2), 170-175. <https://doi.org/10.1038/s41592-020-01056-5>
- De Oliveira, R., Rimbert, H., Balfourier, F., Kitt, J., Dymant, E., Vrána, J., Doležel, J., Cattonaro, F., Paux, E., and Choulet, F. (2020). Structural Variations Affecting Genes and Transposable Elements of Chromosome 3B in Wheats. *Frontiers in Genetics*, **11**, 891. <https://doi.org/10.3389/fgene.2020.00891>

- Diesh, C., Stevens, G.J., Xie, P., De Jesus Martinez, T., Hershberg, E.A., Leung, A., Guo, E., Dider, S., Zhang, J., Bridge, C., Hogue, G., Duncan, A., Morgan, M., Flores, T., Bimber, B.N., Haw, R., Cain, S., Buels, R.M., Stein, L.D., and Holmes, I.H. (2023). JBrowse 2: a modular genome browser with views of synteny and structural variation. *Genome Biology*, **24**(1), 74. <https://doi.org/10.1186/s13059-023-02914-z>
- Droc, G., Larivière, D., Guignon, V., Yahiaoui, N., This, D., Garsmeur, O., Dereeper, A., Hamelin, C., Argout, X., Dufayard, J.-F., Lenglle, J., Baurens, F.-C., Cenci, A., Pitollat, B., D'Hont, A., Ruiz, M., Rouard, M., and Bocs, S. (2013). The Banana Genome Hub. *Database*, **2013**, bat035. <https://doi.org/10.1093/database/bat035>
- Du, Z.-Z., He, J.-B., and Jiao, W.-B. (2025). Plant graph-based pangenomics: techniques, applications, and challenges. *ABIOTECH*, **6**(2), 361-376. <https://doi.org/10.1007/s42994-025-00206-7>
- Dutilh, B.E., and C.P.-G. Consortium. (2016). Computational pan-genomics: status, promises and challenges. *Briefings in Bioinformatics*, **19**(1), 118-135. <https://doi.org/10.1093/bib/bbw089>
- Fuentes, R.R., Chebotarov, D., Duitama, J., Smith, S., De la Hoz, J.F., Mohiyuddin, M., Wing, R. A., McNally, K. L., Tatarinova, T., Grigoriev, A., Mauleon, R., and Alexandrov, N. (2019). Structural variants in 3000 rice genomes. *Genome Research*, **29**(5), 870-880. <https://doi.org/10.1101/gr.241240.118>
- 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., ... and Fei, Z. (2019). The tomato pan-genome uncovers new genes and a rare allele regulating fruit flavor. *Nature Genetics*, **51**(6), 1044-1051. <https://doi.org/10.1038/s41588-019-0410-2>
- Garrison, E., Guarracino, A., Heumos, S., Villani, F., Bao, Z., Tattini, L., ... and Prins, P. (2024). Building pangenome graphs. *Nature Methods*, **21**(11), 2008-2012.
- Glick, L., and Mayrose, I. (2023). The Effect of Methodological Considerations on the Construction of Gene-Based Plant Pangenomes. *Genome Biology and Evolution*, **15**(7), evad121. <https://doi.org/10.1093/gbe/evad121>
- Guarracino, A., Heumos, S., Nahnsen, S., Prins, P., and Garrison, E. (2022). ODGI: understanding pangenome graphs. *Bioinformatics*, **38**(13), 3319-3326. <https://doi.org/10.1093/bioinformatics/btac308>
- Hall, N., Montgomery, J., Chen, J., Saski, C., Matzrafi, M., Westra, P., Gaines, T., and Patterson, E. (2025). *FHY3/FAR1* transposable elements generate adaptive genetic variation in the *Bassia scoparia* genome. *Pest Management Science*, **81**(8), 4393-4402. <https://doi.org/10.1002/ps.8798>
- Hassan, A.H., Mokhtar, M.M., and El Allali, A. (2024). Transposable elements: multifunctional players in the plant genome. *Frontiers in Plant Science*, **14**, 1330127. <https://doi.org/10.3389/fpls.2023.1330127>
- He, Q., Tang, S., Zhi, H., Chen, J., Zhang, J., Liang, H., Alam, O., Li, H., Zhang, H., Xing, L., Li, X., Zhang, W., Wang, H., Shi, J., Du, H., Wu, H., Wang, L., Yang, P., Xing, L., and Diao, X. (2023). A graph-based genome and pan-genome variation of the model plant *Setaria*. *Nature Genetics*, **55**(7), 1232-1242. <https://doi.org/10.1038/s41588-023-01423-w>
- Hickey, G., Monlong, J., Ebler, J., Novak, A.M., Eizenga, J.M., Gao, Y., Human Pangenome Reference Consortium, Marschall, T., Li, H., and Paten, B. (2024). Pangenome graph construction from genome alignments with Minigraph-Cactus. *Nature Biotechnology*, **42**(4), 663-673. <https://doi.org/10.1038/s41587-023-01793-w>
- Hu, G., Cheng, L., Cheng, Y., Mao, W., Qiao, Y., and Lan, Y. (2022). Pan-genome analysis of three main Chinese chestnut varieties. *Frontiers in Plant Science*, **13**, 916550. <https://doi.org/10.3389/fpls.2022.916550>
- Hu, H., Li, R., Zhao, J., Batley, J., and Edwards, D. (2024). Technological Development and Advances for Constructing and Analyzing Plant Pangenomes. *Genome Biology and Evolution*, **16**(4). <https://doi.org/10.1093/gbe/evae081>
- 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**(1), 118. <https://doi.org/10.1038/s41467-024-55260-4>
- Hu, Z., Sun, C., Lu, K., Chu, X., Zhao, Y., Lu, J., Shi, J., and Wei, C. (2017). EUPAN enables pan-genome studies of a large number of eukaryotic genomes. *Bioinformatics*, **33**(15), 2408-2409. <https://doi.org/10.1093/bioinformatics/btx170>
- Jayakodi, M., Schreiber, M., Stein, N. and Mascher, M. (2021). Building pan-genome infrastructures for crop plants and their use in association genetics. *DNA Research*, **28**(1), dsaa030. <https://doi.org/10.1093/dnares/dsaa030>
- Jiang, M., Chen, M., Zeng, J., Du, Z., and Xiao, J. (2024). A comprehensive evaluation of the potential of three next-generation short-read-based plant pan-genome construction strategies for the identification of novel non-reference sequence. *Frontiers in Plant Science*, **15**, 1371222. <https://doi.org/10.3389/fpls.2024.1371222>
- Jiang, M., Qian, Q., Lu, M., Chen, M., Fan, Z., Shang, Y., Bu, C., Du, Z., Song, S., Zeng, J., and Xiao, J. (2025). PlantPan: A comprehensive multi-species plant pan-genome database. *The Plant Journal*, **122**(1). <https://doi.org/10.1111/tpj.70144>
- Jonkheer, E.M., van Workum, D.-J.M., Sheikhzadeh Anari, S., Brankovics, B., de Haan, J. R., Berke, L., van der Lee, T. A. J., de Ridder, D., and Smit, S. (2022). PanTools v3: functional annotation, classification and phylogenomics. *Bioinformatics*, **38**(18), 4403-4405. <https://doi.org/10.1093/bioinformatics/btac506>
- Kolmogorov, M., Yuan, J., Lin, Y., and Pevzner, P. A. (2019). Assembly of long, error-prone reads using repeat graphs. *Nature Biotechnology*, **37**(5), 540-546. <https://doi.org/10.1038/s41587-019-0072-8>
- Koren, S., Walenz, B.P., Berlin, K., Miller, J.R., Bergman, N.H., and Phillippy, A.M. (2017). Canu: scalable and accurate long-read assembly via adaptive *k*-mer weighting and repeat separation. *Genome Research*, **27**(5), 722-736. <https://doi.org/10.1101/gr.215087.116>
- Li, B., Li, X., Wang, Y., Liu, X., Li, K., Li, R., Faiz, M.U., Darlington, U.T., Rao, Z., Wu, Q., Yue, Y., Wang, S., Fei, Z., Liu, Y., Zheng, Y., and Yue, J. (2025). KPGD: A kiwifruit pangenome database for comprehensive mining of genetic diversity in the genus *Actinidia*. *Plant Communications*, **6**(8), 101373. <https://doi.org/10.1016/j.xplc.2025.101373>
- Li, J., Yuan, D., Wang, P., Wang, Q., Sun, M., Liu, Z., Si, H., Xu, Z., Ma, Y., Zhang, B., Pei, L., Tu, L., Zhu, L., Chen, L.-L., Lindsey, K., Zhang, X., Jin, S., and Wang, M. (2021). Cotton pan-genome

- retrieves the lost sequences and genes during domestication and selection. *Genome Biology*, **22**(1), 119. <https://doi.org/10.1186/s13059-021-02351-w>
- Li, Q., Qiao, X., Li, L., Gu, C., Yin, H., Qi, K., Xie, Z., Yang, S., Zhao, Q., Wang, Z., Yang, Y., Pan, J., Li, H., Wang, J., Wang, C., Rieseberg, L. H., Zhang, S., and Tao, S. (2024). Haplotype-resolved T2T genome assemblies and pangenome graph of pear reveal diverse patterns of allele-specific expression and the genomic basis of fruit quality traits. *Plant Communications*, **5**(10), 101000. <https://doi.org/10.1016/j.xplc.2024.101000>
- Liu, H., Wang, X., Liu, S., Huang, Y., Guo, Y.-X., Xie, W.-Z., Liu, H., Tahir ul Qamar, M., Xu, Q., and Chen, L.-L. (2022). Citrus Pan-Genome to Breeding Database (CPBD): A comprehensive genome database for citrus breeding. *Molecular Plant*, **15**(10), 1503-1505. <https://doi.org/10.1016/j.molp.2022.08.006>
- Liu, Z., Wang, N., Su, Y., Long, Q., Peng, Y., Shangguan, L., Zhang, F., Cao, S., Wang, X., Ge, M., Xue, H., Ma, Z., Liu, W., Xu, X., Li, C., Cao, X., Ahmad, B., Su, X., Liu, Y., and Zhou, Y. (2024). Grapevine pangenome facilitates trait genetics and genomic breeding. *Nature Genetics*, **56**(12), 2804-2814. <https://doi.org/10.1038/s41588-024-01967-5>
- Loegler, V., Friedrich, A., and Schacherer, J. (2025). Dynamics of genome evolution in the era of pangenome analysis. *Cell Genomics*, **5**, 101067. <https://doi.org/10.1016/j.xgen.2025.101067>
- 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. <https://doi.org/10.1093/bib/bbae588>
- Montenegro, J.D., Golicz, A.A., Bayer, P.E., Hurgobin, B., Lee, H., Chan, C.K.K., and Edwards, D. (2017). The pan-genome of hexaploid bread wheat. *The Plant Journal*, **90**(5), 1007-1013. <https://doi.org/10.1111/tj.13515>
- Munir, F., Saba, N.U., Arveen, M., Siddiqua, A., Ahmad, J., and Amir, R. (2020). Pan-genomics of plants and its applications. In *Pan-genomics: Applications, Challenges, and Future Prospects* (pp. 285-306). Academic Press.
- Naithani, S., Deng, C.H., Sahu, S.K., and Jaiswal, P. (2023). Exploring Pan-Genomes: An Overview of Resources and Tools for Unraveling Structure, Function, and Evolution of Crop Genes and Genomes. *Biomolecules*, **13**(9), 1403. <https://doi.org/10.3390/biom13091403>
- Oshkin, I. Y., Miroshnikov, K.K., Grouzdev, D.S., and Dedysh, S.N. (2020). Pan-Genome-Based Analysis as a Framework for Demarcating Two Closely Related Methanotroph Genera *Methylocystis* and *Methylosinus*. *Microorganisms*, **8**(5), 768. <https://doi.org/10.3390/microorganisms8050768>
- Qiao, Q., Edger, P.P., Xue, L., Qiong, L., Lu, J., Zhang, Y., Cao, Q., Yocca, A. E., Platts, A. E., Knapp, S. J., Van Montagu, M., Van de Peer, Y., Lei, J., and Zhang, T. (2021). Evolutionary history and pan-genome dynamics of strawberry (*Fragaria* spp.). *Proceedings of the National Academy of Sciences*, **118**(45). <https://doi.org/10.1073/pnas.2105431118>
- Raza, A., Bohra, A., and Varshney, R.K. (2023). Pan-genome for pearl millet that beats the heat. *Trends in Plant Science*, **28**(8), 857-860. <https://doi.org/10.1016/j.tplants.2023.04.016>
- Rijzaani, H., Bayer, P. E., Rouard, M., Doležel, J., Batley, J., and Edwards, D. (2022). The pangenome of banana highlights differences between genera and genomes. *The Plant Genome*, **15**(1), e20100. <https://doi.org/10.1002/tpg2.20100>
- Roy, A., Chandra, T., Mondal, R., Molla, J., Jaiswal, S., Srivastava, M., Kumar, D., Molla, K. A., and Iqbal, M.A. (2025). Genomic resources and genetic improvement of vital tropical and subtropical fruit crops: current status and prospects. *AoB PLANTS*, **17**(6), plaf059. <https://doi.org/10.1093/aobpla/plaf059>
- Ruperao, P., Thirunavukkarasu, N., Gandham, P., Selvanayagam, S., Govindaraj, M., Nebie, B., Manyasa, E., Gupta, R., Das, R.R., Odeny, D.A., Gandhi, H., Edwards, D., Deshpande, S.P., and Rathore, A. (2021). Sorghum Pan-Genome Explores the Functional Utility for Genomic-Assisted Breeding to Accelerate the Genetic Gain. *Frontiers in Plant Science*, **12**, 666342. <https://doi.org/10.3389/fpls.2021.666342>
- Sabety, J., Svara, A., Tegtmeier, R., Feulner, H., Cho, P., Sakina, A., Hickok, D., and Khan, A. (2024). Unlocking diversity from wild relatives of perennial fruit crops in the pan-genomics era. *Current Opinion in Plant Biology*, **82**, 102652. <https://doi.org/10.1016/j.pbi.2024.102652>
- Sarawad, A., Hosagoudar, S., and Parvatikar, P. (2025). Pan-genomics: Insight into the Functional Genome, Applications, Advancements, and Challenges. *Current Genomics*, **26**(1), 2-14. <https://doi.org/10.2174/0113892029311541240627111506>
- Schatz, M.C., Maron, L.G., Stein, J.C., Wences, A.H., Gurtowski, J., Biggers, E., Lee, H., Kramer, M., Antoniou, E., Ghiban, E., Wright, M. H., Chia, J., Ware, D., McCouch, S. R., and McCombie, W. R. (2014). Whole genome *de novo* assemblies of three divergent strains of rice, *Oryza sativa*, document novel gene space of aus and indica. *Genome Biology*, **15**(11), 506. <https://doi.org/10.1186/s13059-014-0506-z>
- She, H., Liu, Z., Xu, Z., Zhang, H., Wu, J., Cheng, F., Wang, X., and Qian, W. (2024). Pan-genome analysis of 13 *Spinacia* accessions reveals structural variations associated with sex chromosome evolution and domestication traits in spinach. *Plant Biotechnology Journal*, **22**(11), 3102-3117. <https://doi.org/10.1111/pbi.14433>
- Sheikhzadeh, S., Schranz, M. E., Akdel, M., de Ridder, D., and Smit, S. (2016). PanTools: representation, storage and exploration of pan-genomic data. *Bioinformatics*, **32**(17), i487-i493. <https://doi.org/10.1093/bioinformatics/btw455>
- Song, J.-M., Guan, Z., Hu, J., Guo, C., Yang, Z., Wang, S., Liu, D., Wang, B., Lu, S., Zhou, R., Xie, W.-Z., Cheng, Y., Zhang, Y., Liu, K., Yang, Q.-Y., Chen, L.-L., and Guo, L. (2020). Eight high-quality genomes reveal pan-genome architecture and ecotype differentiation of *Brassica napus*. *Nature Plants*, **6**(1), 34-45. <https://doi.org/10.1038/s41477-019-0577-7>
- Springer, N. M., Ying, K., Fu, Y., Ji, T., Yeh, C.-T., Jia, Y., Wu, W., Richmond, T., Kitzman, J., Rosenbaum, H., Iniguez, A. L., Barbazuk, W. B., Jeddelloh, J. A., Nettleton, D., and Schnable, P. S. (2009). Maize Inbreds Exhibit High Levels of Copy Number Variation (CNV) and Presence/Absence Variation (PAV) in Genome Content. *PLoS Genetics*, **5**(11), e1000734. <https://doi.org/10.1371/journal.pgen.1000734>
- Stanke, M., Keller, O., Gunduz, I., Hayes, A., Waack, S., and Morgenstern, B. (2006). AUGUSTUS: ab initio prediction of alternative transcripts. *Nucleic Acids Research*, **34**(Web Server), W435-W439. <https://doi.org/10.1093/nar/gkl200>
- Sun, X., Jiao, C., Schwaninger, H., Chao, C.T., Ma, Y., Duan, N., Khan, A., Ban, S., Xu, K., Cheng, L., Zhong, G.-Y., and Fei, Z. (2020). Phased diploid genome assemblies and pan-genomes provide

- insights into the genetic history of apple domestication. *Nature Genetics*, **52**(12), 1423-1432. <https://doi.org/10.1038/s41588-020-00723-9>
- Sun, Y., Wang, J., Li, Y., Jiang, B., Wang, X., Xu, W.-H., Wang, Y.-Q., Zhang, P.-T., Zhang, Y.-J., and Kong, X.-D. (2022). Pan-Genome Analysis Reveals the Abundant Gene Presence/Absence Variations Among Different Varieties of Melon and Their Influence on Traits. *Frontiers in plant science*, **13**, 835496. <https://doi.org/10.3389/fpls.2022.835496>
- Tahir ul Qamar, M., Fatima, K., Rao, M. J., Tang, Q., Sadaqat, M., Ding, B., Chen, L.-L., and Zhu, X.-T. (2025). Comparative genomics profiling of Citrus species reveals the diversity and disease responsiveness of the GLP pangenes family. *BMC Plant Biology*, **25**(1), 388. <https://doi.org/10.1186/s12870-025-06397-x>
- Tettelin, H., Maignani, V., Cieslewicz, M.J., Donati, C., Medini, D., Ward, N.L., Angiuoli, S. V., Crabtree, J., Jones, A.L., Durkin, A. S., DeBoy, R. T., Davidsen, T.M., Mora, M., Scarselli, M., Margarit y Ros, I., Peterson, J. D., Hauser, C. R., Sundaram, J.P., Nelson, W. C., ... and Fraser, C.M. (2005). Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: Implications for the microbial “pan-genome.” *Proceedings of the National Academy of Sciences*, **102**(39), 13950-13955. <https://doi.org/10.1073/pnas.0506758102>
- Tong, C., Jia, Y., Hu, H., Zeng, Z., Chapman, B., and Li, C. (2025). Pangenome and pantranscriptome as the new reference for gene-family characterization: A case study of basic helix-loop-helix (bHLH) genes in barley. *Plant Communications*, **6**(1), 101190. <https://doi.org/10.1016/j.xplc.2024.101190>
- Tonkin-Hill, G., MacAlasdair, N., Ruis, C., Weimann, A., Horesh, G., Lees, J.A., Gladstone, R.A., Lo, S., Beaudoin, C., Floto, R. A., Frost, S.D.W., Corander, J., Bentley, S.D., and Parkhill, J. (2020). Producing polished prokaryotic pangenomes with the Panaroo pipeline. *Genome Biology*, **21**(1), 180. <https://doi.org/10.1186/s13059-020-02090-4>
- Torkamaneh, D., Lemay, M., and Belzile, F. (2021). The pan-genome of the cultivated soybean (PanSoy) reveals an extraordinarily conserved gene content. *Plant Biotechnology Journal*, **19**(9), 1852-1862. <https://doi.org/10.1111/pbi.13600>
- Urhan, A., and Abeel, T. (2021). A comparative study of pan-genome methods for microbial organisms: *Acinetobacter baumannii* pan-genome reveals structural variation in antimicrobial resistance-carrying plasmids. *Microbial Genomics*, **7**(11). <https://doi.org/10.1099/mgen.0.000690>
- Wang, Y., Li, P., Zhu, Y., Zhang, F., Zhang, S., He, Y., Wu, Y., Lin, Y., Wang, H., Ren, W., Wang, L., Yang, Y., Wang, R., Zheng, P., Liu, Y., Wang, S., and Yue, J. (2024). Graph-Based Pangenome of *Actinidia chinensis* Reveals Structural Variations Mediating Fruit Degreening. *Advanced Science*, **11**(28), e2400322. <https://doi.org/10.1002/advs.202400322>
- Weiss-Schneeweiss, H., Emadzade, K., Jang, T.-S., and Schneeweiss, G.M. (2013). Evolutionary consequences, constraints and potential of polyploidy in plants. *Cytogenetic and Genome Research*, **140**(2-4), 137-150. <https://doi.org/10.1159/000351727>
- Yang, S., Cai, J., Wang, M., Liu, W., Yan, J., Jiang, B., and Xie, D. (2023). The construction and analysis of wax gourd pangenome uncover fruit quality-related and resistance genes. *Scientia Horticulturae*, **318**, 112084. <https://doi.org/10.1016/j.scienta.2023.112084>
- Yocca, A.E., Platts, A., Alger, E., Teresi, S., Mengist, M.F., Benevenuto, J., Ferrão, L.F.V, Jacobs, M., Babinski, M., Magallanes-Lundback, M., Bayer, P., Golicz, A., Humann, J. L., Main, D., Espley, R. V, Chagné, D., Albert, N. W., Montanari, S., Vorsa, N., and Edger, P. P. (2023). Blueberry and cranberry pangenomes as a resource for future genetic studies and breeding efforts. *Horticulture Research*, **10**(11), uhad202. <https://doi.org/10.1093/hr/uhad202>
- Yuan, Y., Bayer, P.E., Batley, J., and Edwards, D. (2021). Current status of structural variation studies in plants. *Plant Biotechnology Journal*, **19**(11), 2153-2163. <https://doi.org/10.1111/pbi.13646>
- Zhang, X., Chen, X., Liang, P., and Tang, H. (2018). Cataloging Plant Genome Structural Variations. *Current Issues in Molecular Biology*, 181-194. <https://doi.org/10.21775/cimb.027.181>
- Zhang, Y., Yang, Z., He, Y., Liu, D., Liu, Y., Liang, C., Xie, M., Jia, Y., Ke, Q., Zhou, Y., Cheng, X., Huang, J., Liu, L., Xiang, Y., Raman, H., Kliebenstein, D. J., Liu, S., and Yang, Q.-Y. (2024). Structural variation reshapes population gene expression and trait variation in 2,105 *Brassica napus* accessions. *Nature Genetics*, **56**(11), 2538-2550. <https://doi.org/10.1038/s41588-024-01957-7>
- Zhou, Y., Zhang, Z., Bao, Z., Li, H., Lyu, Y., Zan, Y., Wu, Y., Cheng, L., Fang, Y., Wu, K., Zhang, J., Lyu, H., Lin, T., Gao, Q., Saha, S., Mueller, L., Fei, Z., Städler, T., Xu, S., ... and Huang, S. (2022). Graph pangenome captures missing heritability and empowers tomato breeding. *Nature*, **606**(7914), 527-534. <https://doi.org/10.1038/s41586-022-04808-9>
- Zimin, A.V., Marçais, G., Puiu, D., Roberts, M., Salzberg, S.L., and Yorke, J.A. (2013). The MaSuRCA genome assembler. *Bioinformatics*, **29**(21), 2669-2677. <https://doi.org/10.1093/bioinformatics/btt476>