Molecular Biology and Evolution, 2026, 43, 1–19
https://doi.org/10.1093/molbev/msag087
Published: 22 April 2026
Review

MBE OXFORD

From structural pangenomes to functional panomics in plants

Noé Cochetel [1]*, **Dario Cantu** [1,2]*

[1]Department of Viticulture and Enology, University of California Davis, Davis, CA, USA
[2]Genome Center, University of California Davis, Davis, CA, USA
***Corresponding authors:** Emails: ndcochetel@ucdavis.edu; dacantu@ucdavis.edu.
Associate editor: Flavia Mascagni

Abstract

Understanding how genome structure and gene content evolve within and among plant species requires analytical frameworks that capture the full spectrum of allelic and structural variation. Plant pangenomes have progressed from gene-focused and linear representations to chromosome-scale, haplotype-resolved graphs that preserve allelic and structural diversity across populations. Attention is increasingly focused on the evolutionary processes that generate and maintain pangenome architecture, alongside the adoption of graph-based coordinate systems that reduce reference bias and enable multi-omics integration. Here, we review plant pangenome paradigms, from early homology-based gene sets to haplotype-resolved graph models, and summarize tradeoffs in construction, mapping, and variant analysis. We synthesize patterns of core and variable gene compartments across studies and examine how transposable elements, gene duplication, and selection shape their contrasting evolutionary dynamics. Using panNLRomes and crop domestication as case studies, we illustrate how graph-based frameworks clarify evolutionary and functional signals obscured by single references, including birth-death dynamics at resistance loci and structural variants associated with domestication. Finally, we discuss emerging applications in pantranscriptomics and panepigenomics and outline key methodological and infrastructural challenges.

Keywords pangenome, sequence graph, panomics, panNLRome, pantranscriptomics, panepigenomics

Introduction: Beyond the single reference genome

Plant genomic diversity spans multiple scales, from point mutations and transposable element (TE) insertions to large structural rearrangements and whole-genome duplications that reshape genome size, structure, and gene content (Bennetzen and Wang 2014; Wendel et al. 2016; Soltis and Soltis 2021). Extensive genomic variation is evident even within a single species, where structural and genic configurations arise through diverse mutational mechanisms and are shaped by the interplay of natural selection, demographic history, and genetic drift (Gaut et al. 2018; Exposito-Alonso et al. 2022; Leventhal et al. 2025). At the individual scale, many plant genomes carry deeply diverged haplotypes that maintain high levels of heterozygosity across large genomic regions (Hu et al. 2021; Li et al. 2023a; Chen et al. 2025a). In long-lived perennials and clonally propagated species, somatic mutations and epigenetic drift further amplify divergence among individuals (Vondras et al. 2019; Zhou et al. 2019; Ibañez and Quadrana 2023; Cochetel et al. 2025).

Until recently, nearly all genomic analyses depended on a single reference assembly, in which one haploid representation served as the coordinate system for an entire species. This approach facilitated comparability across studies and provided a common framework for research communities (Kaye and Wasserman 2021). Yet as population-scale sequencing expanded, it became clear that a single reference captures only a fraction of the structural variation within a species. Insertions, deletions, and duplications absent from the reference lead to read misalignment and inaccurate variant calls (Fig. 1a) (Igolkina et al. 2025). These structural variants (SVs) directly affect loci in reference genomes that may lack genes and regulatory sequences present in other individuals (Fig. 1b), obscuring lineage-specific features (Sherman and Salzberg 2020; Lei et al. 2021). Moreover, most references are reduced to a haploid representation, discarding allelic and haplotype diversity (Wong et al. 2020; Della Coletta et al. 2021). Lineage-specific sequences absent from the reference may include regions carrying signatures of recent selection or adaptive divergence (Beaulieu et al. 2025; Feng et al. 2025a). Because evolutionary analyses rely on accurate estimates of allele frequency, linkage, and structural variation, reference bias can distort conclusions about selection, demographic history, and adaptive dynamics (Paten et al. 2017; Sirén et al. 2021; Wu et al. 2024; Akopyan et al. 2025).

Received: December 20, 2025. **Revised:** March 17, 2026. **Accepted:** March 19, 2026

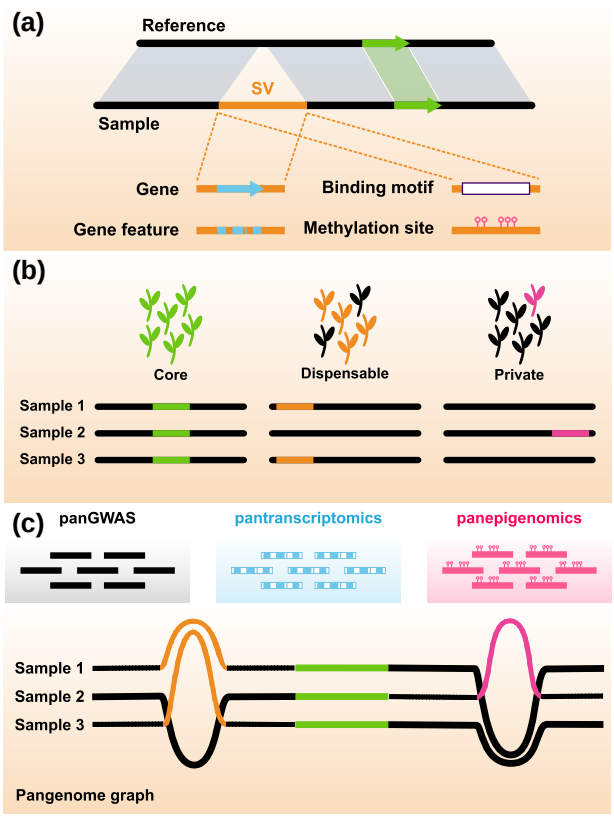

Figure 1 Sequence graphs as a support for panomics. a) Single reference genome fails to capture the genomic diversity of a population. b) Pangenome compartments. The core genome contains sequences/genes present in all individuals, while the dispensable and private genomes are sequences/genes present in some or in a single individual, respectively. c) A sequence graph representing a pangenome can be used to map omics data such as resequencing data, transcriptomics, and epigenomics.

Reference bias propagates through any mapping-dependent analysis. In transcriptomics, reads from paralogous genes or divergent alleles may be misassigned or lost entirely (Fig. 1c) (Stevenson et al. 2013), and genotype-specific references have proven essential for accurate transcript quantification in barley (Guo et al. 2022), rice (Slabaugh et al. 2019), and wheat (Coombes et al. 2024). In epigenomics, methylation sites absent from the reference cannot be detected (Fig. 1c) (Liu et al. 2012; LaBarre et al. 2019; Wulfridge et al. 2019; Zhou et al. 2020), as demonstrated in *Arabidopsis* (Kawakatsu et al. 2016) and grapevine (Cochetel et al. 2025). Similarly, genome-inferred annotations constrain proteomic analysis to reference-encoded proteins, limiting peptide identification, enzyme discovery, and pathway prediction (Mudge and Harrow 2016).

Pangenomics addresses these limitations by integrating multiple genome assemblies into a unified framework that captures both shared and lineage-specific sequences, including SVs, alternative haplotypes, and gene presence–absence variation (PAV) (Golicz et al. 2016a; Sherman and Salzberg 2020). As high-quality assemblies accumulate across species, pangenomes provide a coherent framework for organizing and comparing genomic diversity (Bayer et al. 2020). Graph-based representations are particularly well suited to this task, offering a scalable coordinate system that extends from single genomes to population-level resolution and facilitates the integration of multiple omics layers (Eizenga et al. 2020).

While pangenomes typically represent individuals from a single species, they can be extended to broader taxonomic scales. By integrating multiple species within a genus, super-pangenomes provide a comparative framework for studying genome evolution across species boundaries, including patterns of introgression, lineage-specific gene family dynamics, and the structural basis of species divergence (Li et al. 2023b; Khan et al. 2024; Shi et al. 2024). In plants, super-pangenomes are particularly valuable for incorporating crop wild relatives, which represent the broader diversity from which domesticated gene pools were derived and often harbor adaptive alleles absent from modern cultivars (Khan et al. 2020; Secomandi et al. 2025).

As plant pangenomics matures, two shifts are becoming clear. First, the field is moving beyond enumerating gene content and structural variation toward interpreting the evolutionary mechanisms that create and maintain pangenome architecture, including the roles of TEs, duplication, selection, and demographic history. Second, pangenome graphs are increasingly used as population-level coordinate systems for functional panomics, supporting expression and epigenomic analyses that reduce reference bias and explicitly model structural diversity. In this review, we outline major pangenome paradigms from gene-based catalogs to haplotype-resolved graphs, synthesize what plant pangenomes have revealed about core and variable compartments, and use panNLRomes (the full repertoire of nucleotide-binding leucine-rich repeat (NLR) loci of a pangenome) and domestication studies to illustrate how graph-based frameworks enable analyses that are difficult with single linear references.

Different types of plant pangenomes

Gene-based pangenomes

Gene-based pangenomes were the first approach developed to describe how gene content varies within a clade. They represent the non-redundant set of genes observed across a collection of individuals sampled from that clade. The concept was first introduced in a comparative study of the bacterium *Streptococcus agalactiae* (Tettelin et al. 2005), which showed that no single genome contains the full repertoire of a species but rather a mixture of core genes shared across isolates and variable genes present in only some.

Gene-based pangenomes comprise a core genome with genes present in all individuals and a dispensable (or accessory) genome, consisting of genes not shared by all individuals (Fig. 1b) (Golicz et al. 2020; Lei et al. 2021; Matthews et al. 2024). The term “variable” is sometimes preferred to avoid implying limited biological relevance (Marroni et al. 2014; Lei et al. 2021). Private (or unique) genes are defined as genes found in only one individual and represent a subset of the variable genome (Fig. 1b). Using fixed thresholds is sensitive to technical artifacts; a single annotation or assembly error in any sample is sufficient to incorrectly reassign a true core gene to the dispensable compartment. Such misclassification can be mitigated by evaluating gene presence/absence using frequencies, ranging from core genes present in all individuals to soft-core

(>90-95%), shell (intermediate frequency), and cloud genes detected only rarely (<1%) (Matthews et al. 2024; Loegler et al. 2026).

Gene-based pangenomes can be constructed using two different strategies. Homology-based approaches compare annotated genomes directly, making results sensitive to annotation and assembly errors as well as to user-defined thresholds for sequence identity and coverage (Matthews et al. 2024; Brůna et al. 2026; Loegler et al. 2026) (Table 1). In contrast, map-to-pan approaches align reads from multiple individuals to a representative set of genome sequences and infer gene presence/absence from read coverage, a process that depends strongly on the accuracy and completeness of the reference annotation (Hu et al. 2017; Wang et al. 2018; Gao et al. 2019) (Table 1). Given the limitations of single references discussed above, the map-to-pan strategy tends to underestimate gene presence at highly copy-number-variable loci, such as disease resistance regions, where collapsed assemblies or ambiguous read mapping may lead to false negatives (Tao et al. 2019; Barragan and Weigel 2021).

Gene-based pangenomes remain widely used because they are relatively straightforward to construct. However, they do not capture variation in noncoding regions, including regulatory sequences and repetitive elements, particularly TEs, which are major drivers of eukaryotic genome evolution (Bennetzen and Wang 2014). This limits evolutionary inference, as much adaptive divergence in plants is mediated by regulatory and repetitive sequence variation rather than by changes in coding sequence alone (Lisch 2013; Olsen and Wendel 2013a; Bennetzen and Wang 2014).

Sequence-based linear pangenomes

Sequence-based linear pangenomes extend gene-based models by incorporating noncoding and intergenic sequences alongside annotated genes. This approach was introduced to capture forms of variation that gene-based methods miss, including regulatory sequences, intergenic regions, and repetitive elements (Table 1). In its simplest implementation, a sequence-based linear pangenome is built by performing pairwise whole-genome alignments between complete assemblies; sequences absent from the reference are extracted and concatenated to produce an expanded linear sequence (Matthews et al. 2024). When multiple assemblies are available, this process can be applied iteratively, although a reference genome must still be designated as the initial backbone.

Most approaches to constructing sequence-based pangenomes share two limitations. First, their reliance on a single reference genome introduces reference bias, particularly when SVs or divergent haplotypes are absent from the reference. Second, a linear representation cannot naturally accommodate alternative alleles or complex structural variation, making it difficult to represent regions where multiple structural configurations coexist within a species. Despite these constraints, sequence-based linear pangenomes provided a practical stepping stone between gene-based approaches and graph-based pangenomes (Bayer et al. 2020).

Graph-based pangenomes

Graph-based pangenomes expand on linear approaches by embedding multiple haplotypes within a unified graphical framework. In plants, they are commonly implemented as sequence graphs (Table 1), which reduce redundancy among input genomes by merging identical regions and representing variable segments as nodes connected by edges (Fig. 1c) (Eizenga et al. 2020). Each genome corresponds to a walk through the graph, that is, an ordered series of connected nodes reflecting its linear sequence. In variation graphs, walks are formalized as paths, which record the linear genome sequences used to construct the graph (Garrison et al. 2018). The coordinate system defined by these paths establishes a direct correspondence between positions in the graph and positions in the linear genomes. Sequence variation is represented as bubbles, subgraphs in which paths diverge from a shared node and reconnect downstream, with each path corresponding to an alternative allele. Snarls generalize this concept to accommodate nested and more complex variant structures (Paten et al. 2018).

An alternative graph format, the de Bruijn graph, represents sequences by decomposing them into all overlapping substrings of fixed length k (k-mers), using these as nodes connected by edges encoding consecutive overlaps (Table 1). Stretches of the graph where each node has exactly one predecessor and one successor can be collapsed into longer, unambiguous sequence segments called unitigs, which serve as the practical units of sequence analysis in de Bruijn-based pangenome tools (Eizenga et al. 2020). Colored de Bruijn graphs extend this framework by tagging each node with the sample or samples from which it originates, enabling multi-sample presence/absence comparisons. However, because sequence identity is fragmented into k-mer-sized pieces, long-range haplotype context is not preserved, and SVs spanning repeat-rich regions are difficult to represent faithfully. Although tools such as PanGenie (Ebler et al. 2022) have been developed to mitigate some of these limitations, de Bruijn approaches do not provide the integrative coordinate framework required for panomic inferences. As variation graphs represent both the method of choice in the field and the most suitable framework for the evolutionary and integrative analyses discussed in this review, we focus exclusively on this approach hereafter.

Compared to linear and gene-based representations, graph-based pangenomes place all input haplotypes within a shared coordinate system, allowing alternative alleles, insertions, deletions, and rearrangements to be represented explicitly as parallel paths. This structure enables unbiased read mapping across haplotypes (Fig. 1c) and supports more accurate genotyping in divergent or structurally complex regions, many of which are enriched for loci under balancing selection, local adaptation, or rapid gene family turnover in plants (Igolkina et al. 2025; Teasdale et al. 2025; Huang et al. 2026).

Among the popular tools to build variation graphs, the variation graph toolkit (vg) (Garrison et al. 2018) relies on variant call format (VCF) input. Although vg graphs have been successfully used to represent SVs (Hickey et al. 2020), they rely on reference-based variant calling. As a result, nested or complex variants remain difficult to capture. Minigraph, another popular approach, constructs graphs that represent variants of at least 100 bp in length (Li et al. 2020). While valued for its computational efficiency, minigraph's omission of small variants such as SNPs limits its use as a complete representation. To address these limitations, recent methods leverage multiple genome assemblies to build base-level pangenome graphs (Garrison et al. 2024; Hickey et al. 2024).

Minigraph-Cactus (MC) begins by constructing an SV graph using minigraph (Li et al. 2020; Hickey et al. 2024), followed by alignment of all assemblies to this graph to produce a base-level pangenome (Hickey et al. 2024). In contrast, the PanGenome Graph Builder

Table 1 Summary of the main methods used to build plant pangenomes.

Method	Principle	Advantages	Limits	Common usage	Examples
Homology-based	Annotated genes from multiple assemblies are clustered into orthogroups based on sequence similarity	Computationally tractable; well-supported tools (OrthoFinder, OrthoMCL); intuitive output	Ignores intergenic variation; sensitive to assembly/annotation quality and clustering thresholds; limited syntenic context	Frequently used as a standalone gene classification layer to build gene-based pangenome; often combined with complementary graph-based approaches focused on SVs	Broomcorn millet (Chen et al. 2023a); Potato (Tang et al. 2022); Sorghum (Tao et al. 2021)
Map-to-pan	Reads or contigs from additional accessions are mapped to an iteratively updated pangene reference; unmapped sequences extend the pangenome	Captures genes absent from initial reference; scalable to large accession sets without full assemblies	Reference-guided; may miss highly diverged sequences; mapping quality limited in repeat-rich regions	Can be used standalone to class the pangene set; compatible with further homology-based classification	*Brassica oleracea* (Golicz et al. 2016b); Cotton (Li et al. 2021); Rice (Wang et al. 2018)
Iterative/ stepwise linear	Non-redundant sequence absent from a reference is extracted from additional assemblies and concatenated to build a linear pan-sequence reference	Captures full sequence diversity including intergenic regions; compatible with standard linear genome tools	Linear structure cannot represent overlapping or conflicting haplotypes; reference bias in iterative approaches	Often combined with homology-based classification	*Brassica napus* (Song et al. 2020); Cucumber (Li et al. 2022); Rice (Qin et al. 2021)
Variation graph	Multiple haplotype sequences are embedded in a single directed graph; nodes represent sequences and each genome is a named path; variation is encoded as bubbles	Preserves long-range haplotype context; explicit SV encoding; reference-free read mapping	Computationally intensive; graph complexity increases with divergence; topology-based inference not as intuitive	Combined with independent gene annotations; standalone topology-based gene classification remains rare	Arabidopsis (Igolkina et al. 2025); Barley (Jayakodi et al. 2024); Cannabis (Lynch et al. 2025)
(colored) de Bruijn	Sequences are decomposed into k-mers (fixed-length substrings) used as nodes; colored de Bruijn graphs tag each k-mer with its sample of origin, enabling multi-sample k-mer presence/absence comparisons	Reference-free construction; efficient for large populations; enables k-mer-based GWAS	Long-range context fragmented into k-mer units; SVs in repeat-rich regions difficult to represent; choice of k affects resolution	Rarely used as a standalone plant pangenome method; primarily applied in population-scale k-mer analyses and GWAS	Arabidopsis, tomato, maize (Voichek and Weigel 2020); *Aegilops tauschii* (Gaurav et al. 2022); Wheat (Zhang et al. 2024a)

(PGGB) does not require selecting an initial reference genome; instead, it infers the graph directly from all-vs-all whole genome alignments (Garrison et al. 2024). After aligning input genomes with wfmash (https://github.com/waveygang/wfmash), the graph is inferred using seqwish (Garrison and Guarracino 2023; Garrison et al. 2024) and polished with smoothxg (Garrison et al. 2024).

Pipeline choice depends on the intended application, but regardless of pipeline, pangenome graphs consistently improve read mapping and variant calling (Kopalli et al. 2025). MC graphs are typically preferred for large resequencing projects due to their computational efficiency (Kopalli et al. 2025). PGGB graphs, by contrast, are better suited for detailed analyses of specific regions because of their completeness (Secomandi et al. 2025). Importantly, the shared graphical format enables interoperability across pipelines; for example, a PGGB graph can be simplified and used for read mapping with the vg toolkit (Garrison et al. 2018).

Haplotype-resolved graph-based pangenomes

Haplotype-resolved graph-based pangenomes extend graph representations by preserving full phasing information for each input genome. Although sequence graphs can be used for haplotype inference (Matthews et al. 2024), many plant genomes are now generated as fully phased assemblies (Chin et al. 2016; Zhang et al. 2025a). Accurate phasing is particularly important for highly heterozygous species and has become increasingly achievable at scale with long-read sequencing. Several assembly strategies now produce phased diploid genomes, including trio binning and hifiasm with or without Hi-C integration (Koren et al. 2018; Cheng et al. 2021, 2022). These approaches yield representations that are substantially more accurate than traditional consensus assemblies, which collapse haplotypes and introduce phasing errors (Lian et al. 2025). By preserving full haplotype structure, these approaches enable direct investigation of how recombination, selection, and demographic history shape genome architecture (Lian et al. 2025; Secomandi et al. 2025; Cheng et al. 2025b; Zhang et al. 2025a).

Haplotype-resolved pangenomes enable analyses that were previously difficult or impossible (Garg et al. 2022). At the evolutionary level, their graph structure supports tracking of haplotype block conservation across populations and examination of chromosomal divergence, recombination suppression, and introgression patterns (Lian et al. 2025). They also enable haplotype-specific SV discovery and high-confidence annotation of variant effects on gene function (Lian et al. 2025). At the functional level, accurate representation of allelic gene models facilitates analyses of dominance relationships (Zhang et al. 2021), and allele-specific expression can be quantified by mapping RNA-seq reads directly to a spliced pangenome graph (Sibbesen et al. 2023). Haplotype-aware pangenomes also provide a framework for allele-specific epigenetic analyses, now feasible with long-read sequencing (Fu et al. 2025; Liu and Conesa 2025). Most current tools, including genome assemblers and variant callers, are optimized for diploid assemblies, but graph-based approaches have recently been extended to polyploid genomes (Cheng et al. 2025a).

Biological and evolutionary insights from pangenomes

Patterns of gene content variation across pangenomes

Allelic variation, as well as expansions, contractions, and diversification of paralogous gene families, can strongly influence gene function and regulation, and contribute to key adaptive and agronomic traits (Frary et al. 2000; Ye et al. 2000; Tian et al. 2019). Accurately capturing these dimensions of gene variation is essential for linking pangenome structure to phenotypic diversity and for enabling meaningful functional and evolutionary inference (Jayakodi et al. 2025; Loegler et al. 2026).

Most plant pangenome studies classify genes using categories defined by their frequency across individuals. Unless specified otherwise, we use core to denote genes present in all or most individuals (including soft-core genes, to facilitate comparisons across studies), dispensable for genes present in several but not all individuals, and private for genes specific to a single individual or present at very low frequency. The use of these three terms has been widely recommended to promote consistency in the field (Loegler et al. 2026).

To survey the current state of the field, we compiled gene-based pangenome studies from the literature, yielding a dataset of 57 studies encompassing 38 species and relatives from 16 plant families, spanning basal land plants (Marchantia) to diverse angiosperms, and ranging from major staple cereals and grain legumes to fruit crops, vegetables, trees, and industrial species (Table 2, see Table S1 for the extended version). Most studies were published between 2022 and 2026 reflecting the rapid growth of the field. Panel sizes varied considerably, from as few as 5 accessions to over 1,000 resequenced individuals.

Homology-based approaches, predominantly OrthoFinder, remain the most widely used strategy to construct gene-based pangenomes (Table S1). Gene families or orthology groups (orthogroups) are identified for all the individuals included in the pangenome. Compartments are then defined by categorizing gene clusters into core, dispensable, and private orthogroups, and are usually expressed as the proportion of total orthogroups. Private gene clusters consistently represent the smallest category, usually below 10% of orthogroups, though this fraction can reach 25% to, 33% in rice (Long et al. 2024), kiwifruit (Yu et al. 2025), tea (Chen et al. 2023b), poplar (Shi et al. 2024), and oak (Liang et al. 2025). The pangene set, defined as the non-redundant union of all genes in the pangenome, is consistently larger than the gene content of any single accession and can reach up to ∼2.5 times its size (Hufford et al. 2021).

Reported core genome fractions vary substantially not only across species but also within the same species across studies. Among the six rice pangenomes compiled in Table 2, for instance, core fractions range from ∼10% to ∼60%. This variation arises from three compounding methodological factors. First, the threshold stringency has a direct and deterministic effect on the gene partitioning: studies requiring soft frequency-based thresholds will allow more genes or orthogroups to be considered as core. Second, the choice of computational framework introduces further requirements: orthology-based methods (e.g. OrthoFinder, OrthoMCL), resequencing-based approaches (e.g. map-to-pan, SGSGeneLoss), and graph-based inferences differ fundamentally in how they define gene family membership, handle paralogs, and represent SVs. Third, sampling size and composition have a noticeable impact on the pangenome compartments, large panels, and studies enriched with wild relatives tend to yield lower core fractions (Fig. 2a). For example, in barley, comparing wild to domesticated accessions and increasing sample size reduce the core genome size (Fig. 2a). In rice, including wild relatives in the panel had a significant impact on the core genome size. It is also worth mentioning that comparing accessions sequenced with

Table 2 Gene content variation in plant pangenome studies.

Plant	Economic usage	Family	Orthogroups/Genes (per individual, if available) (C|D|P%)			Sampling (Species | type)	References
Marchantia	Basal lineage/bryophyte	Marchantiaceae	28.4	57.8	13.8	2 | wild	Beaulieu et al. (2025)
…	…	…	35.2	49.2	15.6	1 | wild	Beaulieu et al. (2025)
Barley	Cereal/staple grain	Poaceae	55.7	29.2	15.1	1 | wild	Feng et al. (2025b)
…	…	…	17.5 (64.7)	82 (33.6)	0.5 (1.7)	1 | 17 cultivars, 36 landraces, 23 wild	Jayakodi et al. (2024)
Broomcorn millet	Cereal/staple grain	Poaceae	54.5	37.1	8.4	1 | 24 cultivars, 8 wild	Chen et al. (2023a)
Foxtail millet	Cereal/staple grain	Poaceae	66.7	29.4	3.9	2 | 82 cultivars, 29 wild	He et al. (2023)
Maize	Cereal/staple grain	Poaceae	92.6/55.7	7.4/44.3	…	Multiple | 508 inbred lines, 183 wild	Gui et al. (2022)
…	…	…	66	32	2	1 | 26 inbred lines	Hufford et al. (2021)
Oat	Cereal/staple grain	Poaceae	24	61	15	23 | 14 cultivars, 21 wild	Zhang et al. (2025b)
Pearl millet	Cereal/staple grain	Poaceae	46.6-52.1	39.8-49.9	0.7-8.7	1 | Mixed cultivars, landraces, wild	Yan et al. (2023)
Rice	Cereal/staple grain	Poaceae	20.8	78.9	0.3	21 | 2 cultivars, 19 wild	Fornasiero et al. (2025)
…	…	…	53.2	39.4	7.4	2 | 129 wild, 16 cultivars	Guo et al. (2025a)
…	…	…	9.7 (33.2)	56.8 (62.6)	33.5 (4.2)	16 | 3 cultivars, 14 wild	Long et al. (2024)
…	…	…	42.6	57.4	…	4 | 213 cultivars, 38 wild	Shang et al. (2022)
…	…	…	30.6	47.5	21.9	2 | 33 cultivars	Qin et al. (2021)
…	…	…	53.5-62.1 (63.4-73.5)	≥37.9 (≥26.5)	…	1 | cultivars	Wang et al. (2018)
Sorghum	Cereal/staple grain	Poaceae	36 (58.8)	63.6 (37.9)	0.4 (3.3)	2 | 15 cultivars, 1 wild	Tao et al. (2021)
Wheat	Cereal/staple grain	Poaceae	73	27	…	1 | cultivars	Jiao et al. (2025)
Sugarcane	Sugar/industrial	Poaceae	88.6	11.1	0.3	4 | including 3 cultivars	Huang et al. (2026)
Apple	Fruit/horticultural	Rosaceae	48.9/69.7	49.6/27	1.5/3.3	5 | 9 cultivars, 4 wild	Wang et al. (2023)
…	…	…	81.3-87.3	18.7-12.7	…	3 | 1 cultivar, 2 wild	Sun et al. (2020)
Citrullus	Fruit/horticultural	Cucurbitaceae	51.6	45.9	2.5	7 | Mixed cultivars, landraces, wild	Zhang et al. (2024b)
Citrus	Fruit/horticultural	Rutaceae	35.8	55.4	8.8	18 | cultivars, wild	Huang et al. (2023)
Grape	Fruit/horticultural	Vitaceae	27	71	2	Multiple | 26 cultivars, 46 wild	Guo et al. (2025b)
…	…	…	39.5	56.6	3.9	4 | 12 cultivars, 5 wild, 1 inbred	Liu et al. (2024a)
…	…	…	67.8	24.2	8	9 | wild	Cochetel et al. (2023)
Kiwifruit	Fruit/horticultural	Actinidiaceae	52	47	1	7 | 2 cultivars, 5 wild	Wu et al. (2025)
…	…	…	32.4	34.6	33	8 | 6 cultivars, 9 wild	Yu et al. (2025)
Peach	Fruit/horticultural	Rosaceae	51.8	36.4	11.8	1 | cultivars	Chen et al. (2025b)
…	…	…	80.6	11.5	7.9	5 | 610 improved cultivars, 283 landraces, 48 ornamental, 80 wild	Li et al. (2025a)
Strawberry	Fruit/horticultural	Rosaceae	41.5/42.9	53.6/-	4.9/-	7 | wild	Qiao et al. (2021)
Tea	Leaf/horticultural	Theaceae	29.8	43.9	26.3	1 | cultivars	Chen et al. (2023b)
Arabidopsis	Brassicaceae/model	Brassicaceae	89	7	4	1 | wild inbreds	Igolkina et al. (2025)
…	…	…	65	17	18	1 | wild inbreds	Lian et al. (2024)
…	…	…	80.7	12.6	6.7	1 | wild inbreds	Kang et al. (2023)

(*continued*)

Table 2 Continued

Plant	Economic usage	Family	Orthogroups/Genes (per individual, if available) (C|D|P%)			Sampling (Species | type)	References
Brassica oleracea	Vegetable/horticultural	Brassicaceae	55.1	43.9	1	1 | 10 cultivars, 1 wild	Guo et al. (2024)
…	…	…	34.8 (62)	63.1 (32.9)	2.1 (5.1)	1 | 25 cultivars, 2 wild	Li et al. (2024)
…	…	…	81.3	16.5	2.2	2 | 9 cultivars, 1 wild	Golicz et al. (2016b)
Brassica rapa	Vegetable/horticultural	Brassicaceae	50.2	46.6	3.2	1 | Multiple morphotypes	Ma et al. (2026)
Cucumber	Vegetable/horticultural	Cucurbitaceae	69.5 (80)	30.5 (–)	…	1 | 9 cultivars, 3 wild	Li et al. (2022)
Lettuce	Vegetable/horticultural	Asteraceae	54	43	3	5 | 7 cultivars, 1 landrace, 4 wild	Cao et al. (2025)
Tomato	Vegetable/horticultural	Solanaceae	54	38.4	7.6	11 | 3 cultivars, 10 wild	Li et al. (2023b)
…	…	…	82.2	14.6	3.2	3 | 520 cultivars, 67 wild	Gao et al. (2019)
Medicago	Legume/model	Fabaceae	35	61	4	1 | including 10 cultivars, 11 landraces, 2 wild	He et al. (2025)
Chickpea	Legume/oilseed/protein	Fabaceae	71	25	4	9 | 2 cultivars, 8 wild	Khan et al. (2024)
Common bean	Legume/oilseed/protein	Fabaceae	59	41	…	1 | Mixed domesticated, wild	Cortinovis et al. (2024)
Pea	Legume/oilseed/protein	Fabaceae	50.7	44.3	5	3 | Mixed cultivars, wild	Yang et al. (2022)
Peanut	Legume/oilseed/protein	Fabaceae	44.4	44.3	11.3	4 | 4 improved cultivars, 1 landrace, 3 wild	Zhao et al. (2025)
Soybean	Legume/oilseed/protein	Fabaceae	50.1 (79.8)	49.9 (19.1)	<0.1 (1.1)	2 | 15 cultivars, 9 landraces, 3 wild	Liu et al. (2020)
Potato	Root/tuber	Solanaceae	17.4	63.1	19.5	~60 | Mixed cultivars, landraces, wild	Bozan et al. (2023)
…	…	…	36.7	55.4	7.9	22 | including 20 landraces, 24 wild	Tang et al. (2022)
Brassica napus	Oilseed/industrial	Brassicaceae	56	42	2	1 | cultivars	Song et al. (2020)
Sunflower	Oilseed/industrial	Asteraceae	72.7	21.9	5.4	1 | cultivars	Hübner et al. (2019)
Cannabis	Industrial/fiber	Cannabaceae	78	21	1	1 | cultivars	Lynch et al. (2025)
Cotton	Industrial/fiber	Malvaceae	62.5	35.5	2	1 | 3 cultivars, 2 landraces, 7 primitive lines	Meng et al. (2025)
…	…	…	81.1	4.4	14.5	1 | Mixed landraces, improved lines	Li et al. (2021)
…	…	…	89.8	7.5	2.7	1 | improved lines	Li et al. (2021)
Poplar	Tree/model	Salicaceae	32 (74.1)	31.8 (22.3)	36.2 (3.6)	18 | including 18 wild	Shi et al. (2024)
Oak	Tree/nut	Fagaceae	32.7	35.4	31.9	1 | 22 accessions	Liang et al. (2025)

C|D|P: core, dispensable, and private. In the Orthogroups/Genes column, values represent the proportion of orthogroups in each pangenome compartment. When available, the average percentage of orthogroups per individual is represented between parentheses. When no clustering was performed, values representing proportions of the pangene set are underlined. Sample size is restricted to the accessions used for the gene-based pangenome construction.

different technologies, short and/or long read sequencing, can have a strong impact on the results. In Marchantia, the authors observed that private orthogroups were enriched in genes from long-read genomes revealing a clear technical bias from the sequencing technology (Beaulieu et al. 2025).

In some studies, the core gene fraction exceeds 80%, as reported for sugarcane (Huang et al. 2026), peach (Li et al. 2025a), Arabidopsis (Igolkina et al. 2025), apple (Sun et al. 2020), *Brassica oleracea* (Golicz et al. 2016b), tomato (Gao et al. 2019), and cotton (Li et al. 2021). These high values are most likely due to the method used for gene classification; resequencing-based approaches with a limited number of references tend to output a smaller variable genome (Table S1). Moreover, because of the frequent genome rearrangements observed in plant lineages, the presence of large syntenic regions can be limited, which complicates the homology-based methods and thus, the reliability of the resulting gene-based pangenome compartmentalization (Manzano-Morales et al. 2023). We also observed that gene classification using graph-inferred methods was far more conservative than homology-based strategies (Table S1); the largest core genome size was observed using such a strategy in *Arabidopsis* when compared to homology-based inferences. In the few studies reporting the orthogroups proportions at both pangenome and individual level, the percentage of core clusters was consistently higher in single individuals than in the pangenome (Fig. 2b). This pattern is primarily mathematical: as more accessions are added, each new genome contributes novel dispensable and private gene families that increase the total orthogroup count while leaving the core mostly unchanged, thereby diluting the core fraction at the pangenome level. The practical implication of the dilution effect is that the core fraction is not a fixed species property but shrinks as sampling increases, a saturation behavior first formalized in bacteria (Tettelin et al. 2005) and well documented in plants (Golicz et al. 2020). Some studies also suggest the potential contribution of multi-copy gene families, as paralogs would fall into the same orthogroup while being counted as distinct genes when characterized within a single accession. (Wang et al. 2023).

Although many studies generated both gene-based and sequence-based pangenomes, the two representations are often analyzed separately, with sequence-based pangenomes primarily used to study variants and gene-based pangenomes used to analyze gene diversity. In only a few cases have gene compartments been inferred directly from graph-based pangenomes bypassing the annotation and homology-clustering limits, for example, from graph topology in wild grapes (Cochetel et al. 2023) and from graph-informed annotation reconciliation in *Arabidopsis* (Igolkina et al. 2025). Recently, Huang et al. (2026) proposed an elegant strategy by inferring the gene content directly from graph-based alignments which they successfully applied on a mixed-ploidy sugarcane panel. To our knowledge, topology-based inference was only attempted using graphs constructed with PGGB, the use of other methods such as the reference-guided Minigraph-Cactus might lead to different results, yet it should still overcome homology-based limitations.

Gene characteristics per pangenome compartment

Core genes are typically characterized by longer coding sequences, a higher number of exons, and higher expression levels; they are less affected by nonsynonymous substitutions and generally perform essential cellular functions (Shang et al. 2022; Chen et al. 2023a; Lian et al. 2024). In contrast, variable genes tend to be shorter, exhibit higher nucleotide diversity, and show lower or more context-dependent expression across tissues or environments. They are frequently associated with TEs and often belong to gene families undergoing expansion or contraction (Qiao et al. 2021; Tao et al. 2021; Wu et al. 2025).

Variable genes are often located in TE-rich regions, where insertions and rearrangements promote PAV and structural diversity. In citrus, TE insertions influence gene expression levels, particularly for genes located near long terminal repeat (LTR) retrotransposons and solo LTRs, which are frequently associated with reduced or more variable expression (Huang et al. 2023). Similarly, LTR retrotransposons, mainly from the Gypsy and Copia families, are major drivers of genome size in lettuce, where enrichment of PAVs near TEs indicates a strong link between structural variation and TE activity (Cao et al. 2025). In soybean, nearly 80% of PAVs derive from repetitive DNA, further highlighting the pervasive contribution of TEs to the variable gene compartment (Liu et al. 2020).

Compared with core genes, variable genes are consistently enriched for biotic and abiotic stress response functions, with NBS-LRR disease resistance gene families particularly overrepresented in the dispensable compartment across diverse species (Golicz et al. 2016b; Dolatabadian et al. 2020; Liu et al. 2020; Kang et al. 2023; Long et al. 2024). Variable genes also contribute to phenotypic diversification beyond stress responses; for example, dispensable genes involved in nutrient metabolism underlie flavor differences among *Brassica oleracea* morphotypes (Li et al. 2024).

SVs, including PAVs, copy number variants (CNVs), and larger insertions or deletions, are a primary mechanism by which this functional diversity is generated and maintained (Gabur et al. 2019; Alonge et al. 2020; Göktay et al. 2021). Ubiquitous across genomes, SVs simultaneously affect gene dosage, chromatin state, and the spatial relationship between genes and their regulatory elements; they shape ecological and evolutionary processes at multiple levels of biological organization, from gene expression within cells to ecotype differentiation and speciation across populations (Mérot et al. 2020).

Altogether, the gene patterns inferred from pangenomic studies are consistent with an evolutionary model in which core genes are maintained by purifying selection due to their essential functions, while the variable compartment serves as a reservoir of adaptive potential shaped by TE-mediated structural rearrangement, gene duplication, and diversifying or balancing selection (Bayer et al. 2020; Jayakodi et al. 2025). However, distinguishing adaptive variation from neutral turnover driven by drift remains a challenge, particularly in species with small effective population sizes or recent bottlenecks, where dispensable genes may persist or be lost stochastically rather than through selection (Gaut et al. 2018). As pangenome datasets expand to include more individuals and broader sampling of wild populations, frequency-based analyses of PAVs, analogous to site frequency spectrum approaches for SNPs, may help resolve the relative contributions of selection and drift to pangenome structure. However, realizing this potential will require overcoming substantial technical obstacles, including the higher error rates of PAV genotyping relative to SNPs, the more complex mutational processes underlying

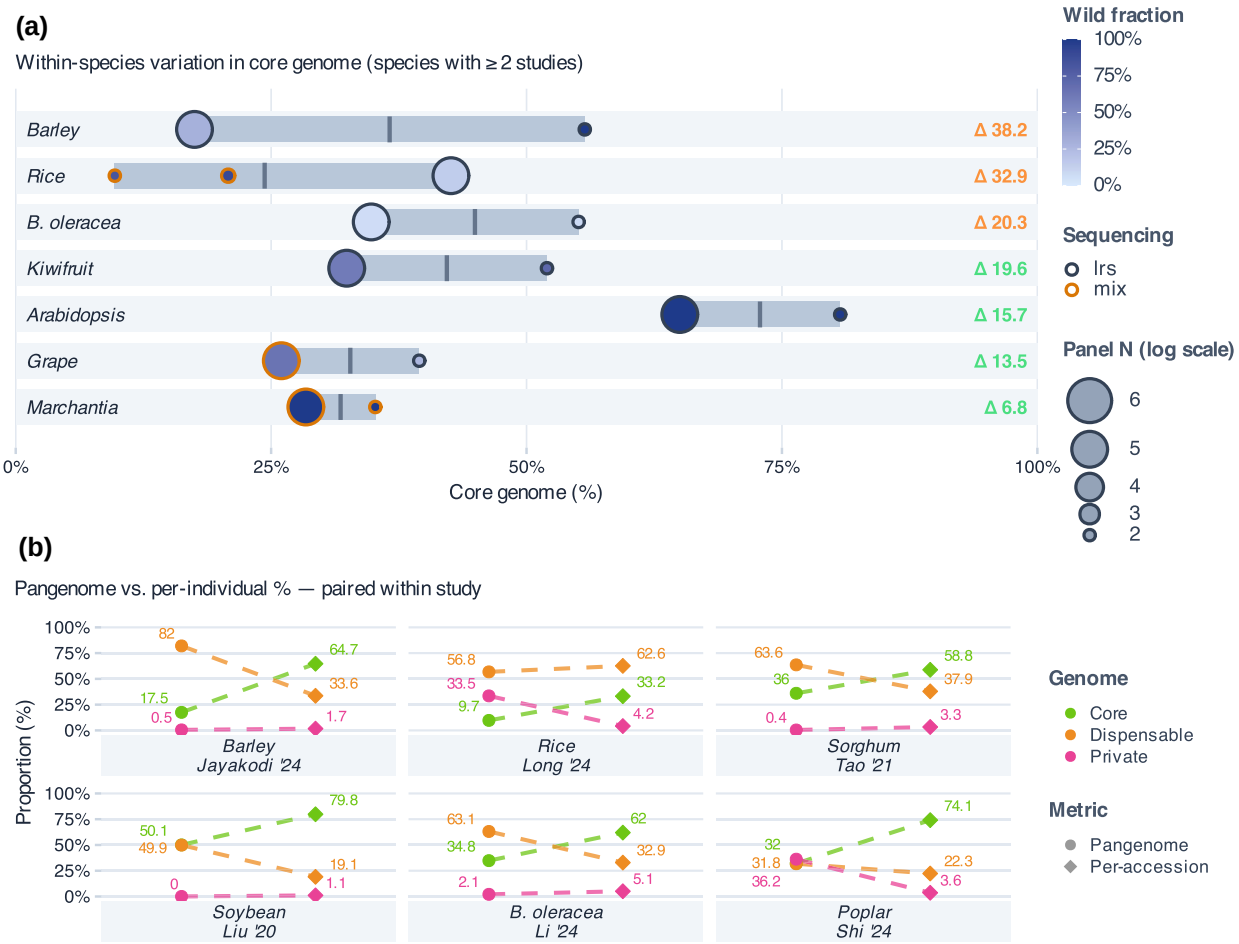

Figure 2 Pangenome compartments comparison. a) Within-species variation in core genome size. For species with at least two homology-based studies, the core genome proportion is represented. Delta values were computed between the minimal and the maximal values for each species. b) Pangenome vs. per-individual pangenome compartment proportions. For each study, the pangenome values are represented as circles on the left, the per-individual values are represented as diamonds on the right.

PAVs, and the fact that population genetic theory developed for point mutations might need to be extended for gene gain and loss dynamics (Gaut et al. 2018; Pokrovac and Pezer 2022).

panNLRomes as a case study of gene family-focused pangenomes

Among gene families, those evolving under strong diversifying or balancing selection are the most structurally dynamic and therefore the most likely to be misrepresented by single references. Pangenomes provide a powerful framework to study the evolution of such families and to identify variants associated with agronomic traits (Schreiber et al. 2024). Disease resistance remains a primary target in plant breeding. Consequently, many pangenomic studies have focused on the NLR loci that encode key components of plant immunity, giving rise to the concept of the panNLRome, the NLR complement of a species pangenome (Barragan and Weigel 2021). Across plant systems, panNLRomes reveal extreme structural and copy-number variability. In the *Solanum* section *Petota*, a pangenome integrating wild and cultivated species found that potato has substantially expanded its NLR repertoire relative to closely related seed-propagated solanaceous crops, consistent with a possible co-evolutionary response to tuber-borne diseases associated with the emergence of vegetative propagation (Tang et al. 2022). A 251-accession rice pangenome graph spanning Asian and African populations showed large variation in NLR copy number at collinear loci across subpopulations (Shang et al. 2022) and the inclusion of wild relatives highlighted their role as major reservoirs of disease resistance (Guo et al. 2025a). In grape populations, a panNLRome revealed an expansion of TIR-NB-LRR genes in North American wild species, implicating this subfamily in downy mildew resistance (Guo et al. 2025b). Similarly, the wheat panNLRome remains unsaturated, reflecting the exceptionally dynamic PAV at NLR loci (Jiao et al. 2025). In contrast, a panNLRome in *Arabidopsis* was found to approach saturation with approximately 40 wild accessions (Van de Weyer et al. 2019). In *Cicer,* a super-pangenome graph enabled systematic cataloging of resistance genes, largely composed of NLRs (Khan et al. 2024). SV genotyping revealed that NLR loci were affected by an average of three SVs per gene, with functional consequences illustrated by the disease resistance gene *RGA1*, where three SVs accumulated during chickpea domestication.

Importantly, panNLRome analyses should not be restricted to gene-centric representations. NLR clusters often span complex

genomic regions (Fig. 3a) that include truncated genes, pseudogenes, and non-expressed fragments that nonetheless contribute to locus evolution (Barragan and Weigel 2021). Recently, Teasdale et al. (2025) introduced a locus-centric strategy for analyzing NLR loci in *Arabidopsis* by defining NLR neighborhoods (Fig. 3b). Rather than solely building a gene atlas and cataloging NLRs, they used pangenome graph traversals to capture both the de novo emergence and complete loss of NLR loci (Fig. 3c) (Teasdale et al. 2025). This locus-centric strategy moves beyond cataloging gene counts and provides a dynamic view of how resistance loci evolve by simultaneously tracking CNVs, domain architecture changes, pseudogenization, and structural rearrangement within their genomic context. Across accessions, the approach reveals how individual loci are born, diversify, degenerate and are entirely lost, processes that are invisible when NLR diversity is reduced to presence/absence scores in a reference-anchored framework. More broadly, panNLRomes illustrate how birth-death evolution, unequal crossing over, and TE-mediated rearrangement interact to generate the rapid structural turnover characteristic of plant defense loci, processes that are likely to operate across other fast-evolving gene families but remain largely unexplored at the pangenome scale (Michelmore and Meyers 1998; Barragan and Weigel 2021).

Wild relatives and pangenomics of crop domestication and adaptive evolution

Crop domestication represents one of the most intensively studied examples of rapid adaptive evolution, characterized by strong directional selection on standing genetic variation over relatively short evolutionary timescales (Olsen and Wendel 2013b; Gaut et al. 2018). Pangenomes are particularly well suited to investigate this process because domestication traits are often associated with SVs, regulatory changes, and gene PAV, classes of variation that are incompletely captured by single-reference genomes (Olsen and Wendel 2013b; Della Coletta et al. 2021). By integrating crop genomes with their wild and landrace relatives, super-pangenomes enable explicit comparison of genome structure across the domestication continuum, revealing which classes of variation were subject to selection and how genome architecture was reshaped during the transition from wild to cultivated forms (Jayakodi et al. 2025).

In broomcorn millet, a pangenome constructed from 32 accessions revealed that TE-derived PAVs contributed to domestication-related traits (Chen et al. 2023a). PAV-based GWAS identified large SVs associated with reduced seed shattering and darker seed color, both targets of directional selection during domestication. In foxtail millet, analysis of 110 globally representative genomes distinguished domestication-related PAVs (wild vs. landrace accessions) from improvement-related PAVs (landraces vs. cultivars), with the greater number of domestication PAVs suggesting stronger selective pressure during the initial domestication phase (He et al. 2023). This study further demonstrated that certain traits were significantly associated with SVs but not with SNPs, highlighting SVs as a major, yet historically undercharacterized, substrate of selection during crop evolution. Similar pangenome-enabled analyses have identified SVs associated with domestication and improvement in tomato (Alonge et al. 2020), cucumber (Li et al. 2022), cotton (Meng et al. 2025), eggplant (Gaccione et al. 2025), and cassava (Xia et al. 2025). In sugarcane, graph-based GWAS was tailored to the complex polyploid genome in DosageGWAS, improving association accuracy and capturing relevant variation missed by conventional methods (Huang et al. 2026). Beyond association mapping, pangenome graphs are beginning to be used to refine population genetic inference more broadly. In *Malus*, the IntervalConnector pipeline projects selective sweep intervals from individual genomes onto a shared coordinate system, enabling cross-accession comparison of selection signals without the reference biases that inflate or obscure sweeps in single-genome analyses (Li et al. 2025b). Together with SV-based GWAS, such approaches move toward a pangenome-native framework for dissecting both the targets and the signatures of selection in crop populations.

Sequence graphs for integrating multi-omics data in panomics

Plant panomics extends pangenomic frameworks beyond genome construction by mapping transcriptomic, epigenomic, and other functional data onto population-level references, enabling reference-aware quantification across individuals and populations (Mishra et al. 2024). The impact of reference choice on functional inference is substantial. In *Arabidopsis*, transcript levels for a subset of genes differed significantly when reads were mapped to individual-specific genomes rather than to the standard TAIR10 reference (Igolkina et al. 2025). Notably, those genes were enriched for CNVs. Methylome analyses were even more sensitive, with non-matching references generating spurious differentially methylated regions (Igolkina et al. 2025), a predictable consequence of the preferential methylation of TEs in plant genomes, whose high structural variability across accessions makes them particularly prone to misalignment when a non-matching reference is used.

Pantranscriptomics

Transcript quantification is one of the areas most directly improved by pangenome-based references. To reduce single-reference bias in RNA-seq analyses, a study in barley implemented a pantranscriptomics strategy in which genotype-specific reference transcript datasets (GsRTDs) were first constructed from representative genomes and then merged into a linear pangenome to generate a pan-RTD (Guo et al. 2025c). Although transcript quantification relied on a conventional linear framework, this approach improved robustness by reducing reference dependence, recovering between 35,500 and 40,800 expressed genes per accession, which improved the overall mapping rates by 11% (Guo et al. 2025c). The authors further note that the respective GsRTDs offer the best performance for RNA-Seq quantification and that the use of a pan-RTD is generally better suited than relying on a common reference. A similar strategy was applied in oat, where RNA-seq reads were mapped to individual genomes and orthology relationships were used to compare gene expression across 23 lines (Avni et al. 2026). Earlier homology-based clustering approaches also reduced reference bias (Contreras-Moreira et al. 2017), but are increasingly being superseded by pangenome-based methods.

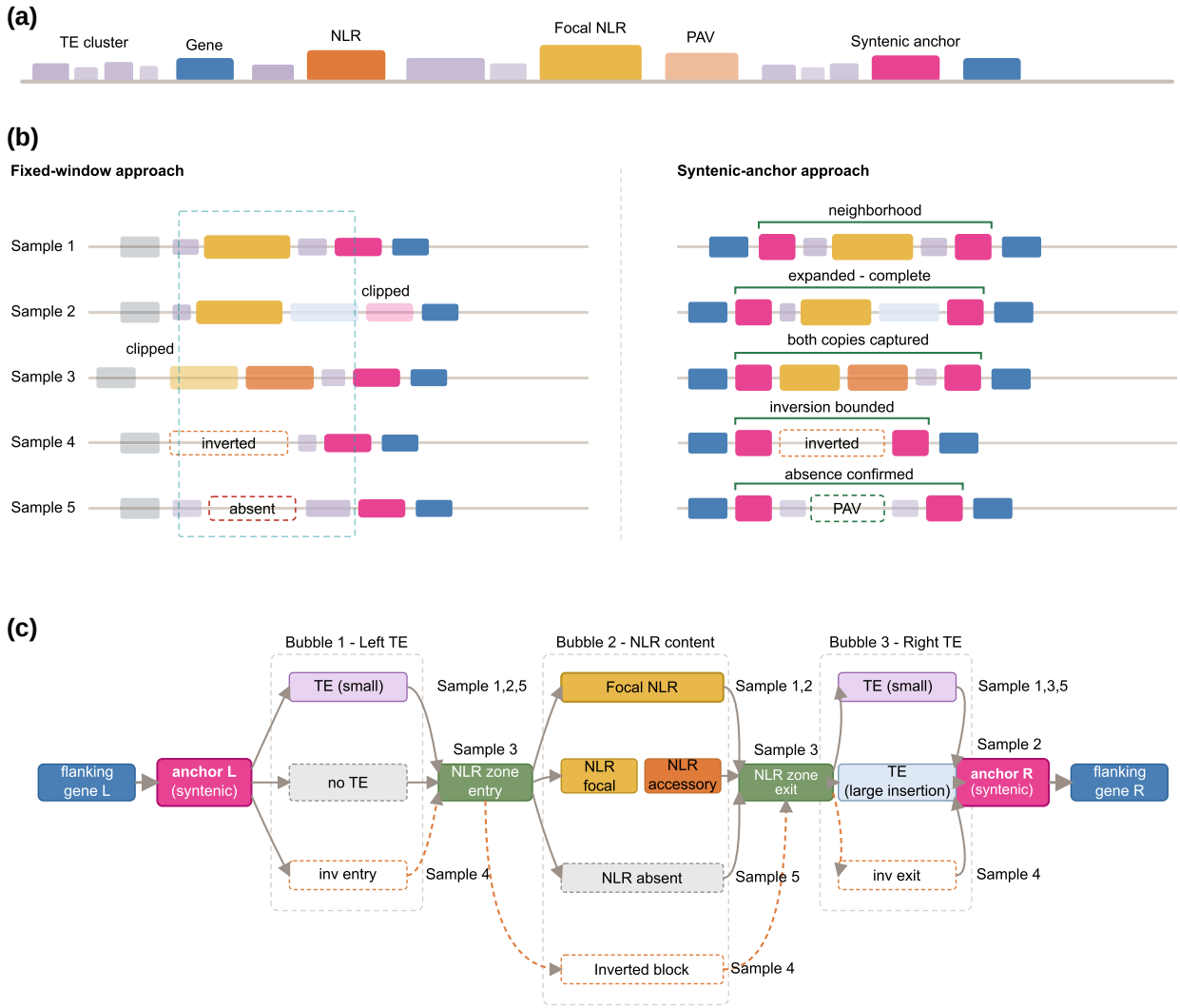

Figure 3 panNLRome representation in sequence graph. a) NLR locus. NLR loci are characterized by a high structural variability, a pronounced TE density and NLR genes often occur in clusters. b) Classic vs. syntenic-anchor approach. In classic approaches, loci comparison is limited to the boundaries of the comparison window defined by alignments. Using syntenic anchors as described by Teasdale et al. (2025), NLR neighborhoods more accurately capture comparable NLR loci across accessions. c) In a pangenome graph, the complexity of NLR neighborhoods is compacted into variation bubbles extending between the two syntenic anchors.

Several tools now enable direct quantification of gene expression on pangenomes, including HISAT2 (Kim et al. 2019), GraphAligner (Rautiainen and Marschall 2020), and RPVG (Sibbesen et al. 2023). Recent developments support haplotype-specific transcript quantification by incorporating gene structural annotations into pangenome graphs, yielding spliced graph representations suitable for transcript-level analyses (Sibbesen et al. 2023). Consistently, the splice-aware graph mapper VG MPMAP improves read mapping and reduces bias in allele-specific expression estimates (Hickey et al. 2020; Sibbesen et al. 2023). In *Brassica napus*, graph-based expression quantification outperformed linear approaches in both gene expression and expression Quantitiative Trait Locus (eQTL) analyses, particularly for transcripts enriched in sequence variation, and most closely matched simulated ground-truth expression levels (Yildiz et al. 2025). Among the transcripts whose expression was associated with SVs, approximately a third showed no association with SNPs, representing a class of SV-driven expression variation entirely invisible to conventional SNP-based eQTL analyses (Yildiz et al. 2025). Collectively, these studies demonstrate that graph-based transcriptomics improve transcript abundance quantification particularly for genes affected by SVs, precisely the gene category that is most poorly represented by single references.

Panepigenomics

Following the development of pantranscriptomics, attention has increasingly turned to other omics layers that are strongly affected by reference genome choice, particularly epigenomics. Although panepigenomics remains at an early stage, initial efforts, mainly in non-plant systems, have begun to emerge. In humans, the use of pangenome references was shown to improve DNA methylation analyses (Dong et al. 2025a). In cattle, MacPhillamy et al. (2024) demonstrated that reference genome choice strongly influences methylation profiling by developing a pairwise coordinate conversion framework based on Liftoff (Shumate and Salzberg 2021) to

enable comparisons between the Angus and Brahman genomes (MacPhillamy et al. 2024). While effective for pairwise comparisons, such approaches scale poorly as additional genomes are incorporated and would benefit from a unified pangenome reference.

In plants, the first application of a graph-based framework to methylation analysis was reported in grapevine (Cochetel et al. 2025). Methylation data were mapped to individual genomes and subsequently projected onto a sequence graph to relate regions and perform differential methylation analysis. This work demonstrated the substantial bias introduced by single-reference genomes, particularly in highly heterozygous species where haplotype resolution is critical (Cochetel et al. 2025). Nevertheless, these approaches remain indirect, as they rely on coordinate projections rather than direct mapping to a pangenome graph.

To date, only one tool has been proposed for direct DNA methylation calling on genome graphs (Zhang et al. 2025c). MethylGrapher enables methylation analysis in CG and non-CG contexts by mapping bisulfite sequencing data directly to pangenome graphs using vg giraffe and projecting graph coordinates to linear space via surjection (Sirén et al. 2021; Zhang et al. 2025c). Advances in sequencing technologies further support the development of panepigenomics. Long-read platforms from Pacific Biosciences and Oxford Nanopore Technologies now enable direct detection of DNA methylation from native reads (Liu and Conesa 2025). Because these reads can be used both for genome assembly and methylation profiling, they reduce biases associated with short-read approaches. Continued reductions in sequencing costs, together with improvements in long-read graph mappers such as GraphAligner (Rautiainen and Marschall 2020) and updated versions of vg giraffe optimized for long reads (Chang et al. 2025), are expected to accelerate the adoption of pangenome-based epigenomic analyses.

Local sequence graphs for targeted analysis of genomic loci

While pangenome graphs typically represent complete chromosome sets, sequence graphs can also be applied at a local scale to interrogate specific, highly complex genomic regions. In the draft human pangenome reference, local graphs were constructed for five clinically relevant loci, capturing known haplotypes and revealing additional previously uncharacterized variants (Liao et al. 2023). Similarly, a PGGB graph built from 94 haplotypes was used to study the amylase locus, enabling haplotype deconvolution from short-read data in regions that were previously difficult to resolve (Bolognini et al. 2024).

In plants, local sequence graphs have been used to characterize complex loci across multiple assemblies. A targeted graph constructed from 76 genome assemblies enabled detailed analysis of the *amy1_1* locus encoding α-amylases (Jayakodi et al. 2024). In grapes, local graphs spanning the powdery mildew resistance loci *Ren6* and *Ren7* revealed extensive haplotypic diversity and private sequence nodes within candidate NLR genes; without haplotype resolution, the NLR gene content and the structural variations specific to the allele providing the resistance would be confounded with the alternative allele, preventing the clear identification of disease resistance sources (Massonnet et al. 2025). Given the repetitive and structurally complex nature of resistance loci, such analyses are difficult using linear references alone. In sorghum, a pangenome-based approach was applied to the *RMES1* locus to identify candidate NLR genes, and the region was subsequently evaluated within a Poaceae super-pangenome to place these variants in an evolutionary context (VanGessel et al. 2025). Locus-focused analyses have long been central to candidate gene discovery, and local pangenome subgraphs now provide the resolution these analyses require.

Conclusions and future directions

Although pangenomics has been widely adopted, constructing and analyzing sequence graphs remain computationally demanding. Graph construction strategies are often constrained by available computational infrastructure, and the choice of tools for read mapping and variant calling is frequently driven by feasibility rather than by graph completeness or biological resolution (Secomandi et al. 2025). Pangenome graphs are typically stored in the graphical fragment assembly (GFA) format, which efficiently captures graph structure and supports community sharing (Eizenga et al. 2020). However, most analysis tools developed for linear references are not directly compatible with GFA and require intermediate conversions. A common workaround is to decompose graphs into sets of variants stored in the VCF (Paten et al. 2018; Liao et al. 2023). While powerful, this transformation can be lossy: nested bubbles may be decomposed into multiple independent variant sites, and multiallelic records generated during decomposition are often incompatible with standard population genomics workflows (Loegler et al. 2026).

Many routine pipelines developed for linear reference genomes have yet to be fully adapted for pangenomes. Several pangenome browsers have been proposed (Pedersen et al. 2017; Durant et al. 2021; Liu et al. 2024b; Miao and Yue 2025), but no broadly adopted standard has emerged. Although synteny viewers can link annotations across genomes (Diesh et al. 2023), defining synteny anchors directly within pangenome graphs remains challenging, particularly because splicing information across multiple genomes is distributed across complex, often nested, graph structures. As a result, synteny analyses are frequently performed using separate gene-centric approaches, which undermines the advantages of graph-based representations, especially for reproducibility and data sharing.

Moreover, public data repositories for pangenomes remain largely unstandardized. Dedicated submission routes for pangenome graphs through established archives such as NCBI and EMBL-EBI have yet to be implemented. The development of centralized repositories, or at a minimum the adoption of standardized formats for data submission and exchange, would substantially improve data accessibility, interoperability, and reuse within the community. In the absence of such infrastructure, authors frequently develop project-specific websites and databases to disseminate their results. Examples include the Solanaceae pangenome database (http://www.bioinformaticslab.cn/SolPGD/#/)(Zhang et al. 2025d), BryoGenomes (http://www.bryogenomes.org/) (Dong et al. 2025b), BnaOmics (https://bnaomics.ocri-genomics.net/)(Cui et al. 2023), SoyOmics (https://ngdc.cncb.ac.cn/soyomics/index)(Liu et al. 2023), GrainGenes (https://graingenes.org/GG3/)(Yao et al. 2022), SoyBase (https://www.soybase.org/)(Brown et al. 2021), RPAN (https://cgm.sjtu.edu.cn/3kricedb/)(Sun et al. 2017), PanBARLEX (https://panbarlex.ipk-gatersleben.de/)(Jayakodi et al. 2024). In addition, Gramene provides

Box 1 Glossary

Accessory genes/genome: Genes or sequences present in some but not all individuals of a pangenome, used interchangeably with **dispensable/variable**.
Bubbles: Representation of structural variations within a graph. Two or more paths diverge from a shared node and reconverge at a following node.
Cloud genes/genome: On a continuous scale, genes or sequences present in only a few individuals of a pangenome.
Core genes/genome: Genes or sequences present in all the individuals of a pangenome.
Dispensable genes/genome: Genes or sequences present in some but not all individuals of a pangenome, used interchangeably with **accessory/variable**.
Gene-based pangenome: Pangenome constructed using genes only.
Haplotypes: A set of genetic variants (alleles) inherited together on the same chromosome. In diploid organisms, each individual has two haplotypes.
Linear reference genome: A single linear genome used as the representative sequence of a species, traditionally termed a reference genome.
Local/sub-graph: A selected genomic region or locus extracted from a pangenome graph.
Nested variant/bubble: Complex bubble structure involving the presence of a smaller variant/bubble embedded inside a larger one.
Nodes: Fundamental unit of the graph. In a sequence graph, nodes represent genomic sequences ranging from a single nucleotide to several base pairs.
Pangenome: In the biological sense, a pangenome is a conceptual term to represent the non-redundant catalog of genome sequences of a species or higher taxonomy levels. To estimate the biological pangenome, we construct pangenomes through computation using sequencing data (Matthews et al. 2024). Unless specified otherwise, the term pangenome is used throughout this review to describe *in silico* pangenomes.
Pangenome graph: Representation of multiple individual genomes in a sequence graph structure.
Paths: Sequences of nodes within a graph representing the genomes of the individuals used to construct the pangenome.
Phasing: Separating haplotypes of an individual.
Presence–absence variant (PAV): Structural variant consisting of the presence or absence of a gene or a genomic region in one individual compared to another, usually to a reference genome.
Private genes/genome: Genes or sequences present in only a single individual of a pangenome. It usually represents an exclusive subset of the **accessory/dispensable/variable** genes/genome.
Sequence graph: Data structure used to store a pangenome as a graph with the nodes representing DNA sequences.
Shell genes/genome: On a continuous scale, genes or sequences present in more than a few (**cloud**) but less than most (**soft-core**) individuals of a pangenome.
Snarls: A generalization of the bubble concept that accommodates nested and more complex variant structures, including variants embedded within other variants.
Splicing graph: Sequence graph representing the splicing junctions of a genome annotation.
Soft-core genes/genome: Genes or sequences present in most of the individuals of a pangenome.
Structural variant (SV): Variation of genomic sequence longer than 50 bp.
Super-pangenome: Pangenome including more than one species, usually combining a cultivar and its wild relatives.
Variable genes/genome: Genes or sequences present in some but not all individuals of a pangenome, used interchangeably with **accessory/dispensable**.

curated access to several crop pangenomes through dedicated portals (https://www.gramene.org/)(Olson et al. 2026).

Advances in long-read sequencing technologies are reshaping pangenome study design. A tiered strategy that combines a limited number of high-quality assemblies with short-read genotyping of larger populations remains cost effective (Schreiber et al. 2024), but continued reductions in long-read sequencing costs may increasingly favor long-read-based genotyping approaches.

Beyond species-level pangenomes, recent studies have begun to explore pangenomics at higher taxonomic scales. A large-scale analysis of 123 bryophyte genomes revealed diversification patterns driven largely by de novo gene origination and microbial gene acquisition, contrasting with trends observed in vascular plants (Dong et al. 2025b). Gene-based pangenomes have similarly provided insights into legume diversification (Wang et al. 2025) and the evolution of floral regulatory genes in *Solanaceae* (Zhang et al. 2025d).

These studies illustrate that pangenomes are increasingly used not only as improved references for variant calling but as tools for evolutionary inference, enabling analyses of gene birth-death dynamics, introgression, and adaptive divergence that were previously inaccessible (Lian et al. 2025; Teasdale et al. 2025). However, the evolutionary potential of pangenome data remains underexploited. Few studies have systematically applied population genetic frameworks to structural variation, for example by analyzing PAV frequency spectra to distinguish selection from drift, or by using graph topology to infer ancestral versus derived structural states (Loegler et al. 2026). Integrating pangenome graphs with phylogenomic methods could further enable reconstruction of reticulate evolutionary histories, particularly in genera where hybridization and introgression are pervasive (Morales-Cruz et al. 2021; Liang et al. 2025).

Collectively, these developments position pangenomics as a central framework for integrating multi-omics data and advancing both comparative and evolutionary genomics. Graph-based pangenomes are beginning to reshape how we study adaptation, speciation, and genome evolution in plants, while also facilitating the discovery of novel breeding targets and enabling population-scale analyses (Mishra et al. 2024; Loegler et al. 2026). Moving beyond linear reference paradigms will be essential for future progress, although widespread adoption of graph-based frameworks will require continued methodological and infrastructural development.

Supplementary material

Supplementary material is available at *Molecular Biology and Evolution* online.

Funding

This work was partially supported by the National Science Foundation grant #1741627 and the National Institute of Food and Agriculture at the U.S. Department of Agriculture NIFA Award # 2022-51181-38240, the E.&J. Gallo Winery, and the Ray Rossi Endowment in Viticulture and Enology.

Conflicts of Interest

None declared.

Data availability

There is no new data associated with this article.

References

Akopyan M, Genchev M, Armstrong EE, Mooney JA. Reference genome choice compromises population genetic analyses. *Cell*. 2025:188:6939–6952.e11. https://doi.org/10.1016/j.cell.2025.08.034.

Alonge M *et al.* Major impacts of widespread structural variation on gene expression and crop improvement in tomato. *Cell*. 2020:182:145–161.e23. https://doi.org/10.1016/j.cell.2020.05.021.

Avni R *et al.* A pangenome and pantranscriptome of hexaploid oat. *Nature*. 2026:649:131–139. https://doi.org/10.1038/s41586-025-09676-7.

Barragan AC, Weigel D. Plant NLR diversity: the known unknowns of pan-NLRomes. *Plant Cell*. 2021:33:814–831. https://doi.org/10.1093/plcell/koaa002.

Bayer PE, Golicz AA, Scheben A, Batley J, Edwards D. Plant pangenomes are the new reference. *Nat Plants.* 2020:6:914–920. https://doi.org/10.1038/s41477-020-0733-0.

Beaulieu C *et al.* The *Marchantia polymorpha* pangenome reveals ancient mechanisms of plant adaptation to the environment. *Nat Genet.* 2025:57:729–740. https://doi.org/10.1038/s41588-024-02071-4.

Bennetzen JL, Wang H. The contributions of transposable elements to the structure, function, and evolution of plant genomes. *Annu Rev Plant Biol.* 2014:65:505–530. https://doi.org/10.1146/annurev-arplant-050213-035811.

Bolognini D *et al.* Recurrent evolution and selection shape structural diversity at the amylase locus. *Nature*. 2024:634:617–625. https://doi.org/10.1038/s41586-024-07911-1.

Bozan I *et al.* Pangenome analyses reveal impact of transposable elements and ploidy on the evolution of potato species. *Proc Natl Acad Sci U S A.* 2023:120:e2211117120. https://doi.org/10.1073/pnas.2211117120.

Brown AV *et al.* A new decade and new data at SoyBase, the USDA-ARS soybean genetics and genomics database. *Nucleic Acids Res*. 2021:49:D1496–D1501. https://doi.org/10.1093/nar/gkaa1107.

Brůna T, Sreedasyam A, Harder AM, Lovell JT. Evolutionary and methodological considerations when interpreting gene presence–absence variation in pangenomes. *NAR Genom Bioinform*. 2026:8:lqag011. https://doi.org/10.1093/nargab/lqag011.

Cao S, Sawettalake N, Shen L. Lactuca super-pangenome provides insights into lettuce genome evolution and domestication. *Nat Commun.* 2025:16:7257. https://doi.org/10.1038/s41467-025-62641-w.

Chang X *et al.* Rapid, accurate long- and short-read mapping to large pangenome graphs with vg Giraffe [preprint]. bioRxiv 678807. 2025. https://doi.org/10.1101/2025.09.29.678807.

Chen J *et al.* Pangenome analysis reveals genomic variations associated with domestication traits in broomcorn millet. *Nat Genet.* 2023a:55:2243–2254. https://doi.org/10.1038/s41588-023-01571-z.

Chen M *et al.* Genomic insights into genome-wide heterozygosity and its impact on walnut adaptive evolution and improvement. *Mol Breed.* 2025a:45:50. https://doi.org/10.1007/s11032-025-01572-2.

Chen S *et al.* Gene mining and genomics-assisted breeding empowered by the pangenome of tea plant *Camellia sinensis*. *Nat Plants.* 2023b:9:1986–1999. https://doi.org/10.1038/s41477-023-01565-z.

Chen W *et al.* Graph pangenome reveals the regulation of malate content in blood-fleshed peach by NAC transcription factors. *Genome Biol.* 2025b:26:7. https://doi.org/10.1186/s13059-024-03470-w.

Cheng H *et al.* Haplotype-resolved assembly of diploid genomes without parental data. *Nat Biotechnol.* 2022:40:1332–1335. https://doi.org/10.1038/s41587-022-01261-x.

Cheng H, Concepcion GT, Feng X, Zhang H, Li H. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nat Methods*. 2021:18:170–175. https://doi.org/10.1038/s41592-020-01056-5.

Cheng L *et al.* Genome analyses and breeding of polyploid crops. *Nat Plants.* 2025a:11:1714–1728. https://doi.org/10.1038/s41477-025-02088-5.

Cheng L *et al.* Leveraging a phased pangenome for haplotype design of hybrid potato. *Nature*. 2025b:640:408–417. https://doi.org/10.1038/s41586-024-08476-9.

Chin C-S *et al.* Phased diploid genome assembly with single-molecule real-time sequencing. *Nat Methods*. 2016:13:1050–1054. https://doi.org/10.1038/nmeth.4035.

Cochetel N *et al.* A super-pangenome of the North American wild grape species. *Genome Biol.* 2023:24:290. https://doi.org/10.1186/s13059-023-03133-2.

Cochetel N *et al.* Phased epigenomics and methylation inheritance in a historical *Vitis vinifera* hybrid. *Genome Biol.* 2025:26:392. https://doi.org/10.1186/s13059-025-03858-2.

Contreras-Moreira B *et al.* Analysis of plant pan-genomes and transcriptomes with GET_HOMOLOGUES-EST, a clustering solution for sequences of the same species. *Front Plant Sci.* 2017:8:184. https://doi.org/10.3389/fpls.2017.00184.

Coombes B, Lux T, Akhunov E, Hall A. Introgressions lead to reference bias in wheat RNA-Seq analysis. *BMC Biol.* 2024:22:56. https://doi.org/10.1186/s12915-024-01853-w.

Cortinovis G *et al.* Adaptive gene loss in the common bean pangenome during range expansion and domestication. *Nat Commun.* 2024:15:6698. https://doi.org/10.1038/s41467-024-51032-2.

Cui X *et al.* BnaOmics: a comprehensive platform combining pangenome and multi-omics data from *Brassica napus*. *Plant Commun.* 2023:4:100609. https://doi.org/10.1016/j.xplc.2023.100609.

Della Coletta R, Qiu Y, Ou S, Hufford MB, Hirsch CN. How the pangenome is changing crop genomics and improvement. *Genome Biol*. 2021:22:3. https://doi.org/10.1186/s13059-020-02224-8.

Diesh C *et al.* JBrowse 2: a modular genome browser with views of synteny and structural variation. *Genome Biol.* 2023:24:74. https://doi.org/10.1186/s13059-023-02914-z.

Dolatabadian A *et al.* Characterization of disease resistance genes in the *Brassica napus* pangenome reveals significant structural variation. *Plant Biotechnol J.* 2020:18:969–982. https://doi.org/10.1111/pbi.13262.

Dong S, *et al.* Bryophytes hold a larger gene family space than vascular plants. *Nat Genet*. 2025b:57:2562–2569. https://doi.org/10.1038/s41588-025-02325-9.

Dong Z, *et al.* Complete reference genome and pangenome improve genome-wide detection and interpretation of DNA methylation using sequencing and array data. *Cell Rep*. 2025a:44:115755. https://doi.org/10.1016/j.celrep.2025.115755.

Durant É, Sabot F, Conte M, Rouard M. Panache: a web browser-based viewer for linearized pangenomes. *Bioinformatics*. 2021:37:4556–4558. https://doi.org/10.1093/bioinformatics/btab688.

Ebler J *et al.* Pangenome-based genome inference allows efficient and accurate genotyping across a wide spectrum of variant classes. *Nat Genet*. 2022:54:518–525. https://doi.org/10.1038/s41588-022-01043-w.

Eizenga JM *et al.* Pangenome graphs. *Annu Rev Genomics Hum Genet.* 2020:21:139–162. https://doi.org/10.1146/annurev-genom-120219-080406.

Exposito-Alonso M *et al.* Genetic diversity loss in the *Anthropocene*. *Science*. 2022:377:1431–1435. https://doi.org/10.1126/science.abn5642.

Feng C, Yang L, Wang J, Shi X, Kang M. *Primulina* pan-genome reveals differential gene retention following whole-genome duplications and provides insights into edaphic specialization. *Cell Rep.* 2025a:44:116763. https://doi.org/10.1016/j.celrep.2025.116763.

Feng J-W *et al.* A haplotype-resolved pangenome of the barley wild relative *Hordeum bulbosum*. *Nature*. 2025b:645:429–438. https://doi.org/10.1038/s41586-025-09270-x.

Fornasiero A *et al.* Oryza genome evolution through a tetraploid lens. *Nat Genet.* 2025:57:1287–1297. https://doi.org/10.1038/s41588-025-02183-5.

Frary A *et al.* Fw2.2: a quantitative trait locus key to the evolution of tomato fruit size. *Science*. 2000:289:85–88. https://doi.org/10.1126/science.289.5476.85.

Fu Y, Timp W, Sedlazeck FJ. Computational analysis of DNA methylation from long-read sequencing. *Nat Rev Genet.* 2025:26:620–634. https://doi.org/10.1038/s41576-025-00822-5.

Gabur I, Chawla HS, Snowdon RJ, Parkin IAP. Connecting genome structural variation with complex traits in crop plants. *Theor Appl Genet.* 2019:132:733–750. https://doi.org/10.1007/s00122-018-3233-0.

Gaccione L *et al.* Graph-based pangenomes and pan-phenome provide a cornerstone for eggplant biology and breeding. *Nat Commun.* 2025:16:9919. https://doi.org/10.1038/s41467-025-64866-1.

Gao L *et al.* The tomato pan-genome uncovers new genes and a rare allele regulating fruit flavor. *Nat Genet*. 2019:51:1044–1051. https://doi.org/10.1038/s41588-019-0410-2.

Garg S, Balboa R, Kuja J. Chromosome-scale haplotype-resolved pangenomics. *Trends Genet.* 2022:38:1103–1107. https://doi.org/10.1016/j.tig.2022.06.011.

Garrison E *et al.* Variation graph toolkit improves read mapping by representing genetic variation in the reference. *Nat Biotechnol.* 2018:36:875–879. https://doi.org/10.1038/nbt.4227.

Garrison E *et al.* Building pangenome graphs. *Nat Methods.* 2024:21:2008–2012. https://doi.org/10.1038/s41592-024-02430-3.

Garrison E, Guarracino A. Unbiased pangenome graphs. *Bioinformatics*. 2023:39:btac743. https://doi.org/10.1093/bioinformatics/btac743.

Gaurav K *et al.* Population genomic analysis of *Aegilops tauschii* identifies targets for bread wheat improvement. *Nat Biotechnol.* 2022:40:422–431. https://doi.org/10.1038/s41587-021-01058-4.

Gaut BS, Seymour DK, Liu Q, Zhou Y. Demography and its effects on genomic variation in crop domestication. *Nat Plants.* 2018:4:512–520. https://doi.org/10.1038/s41477-018-0210-1.

Göktay M, Fulgione A, Hancock AM. A new catalog of structural variants in 1,301 *A. thaliana* lines from Africa, Eurasia, and North America reveals a signature of balancing selection at defense response genes. *Mol Biol Evol.* 2021:38:1498–1511. https://doi.org/10.1093/molbev/msaa309.

Golicz AA *et al.* The pangenome of an agronomically important crop plant *Brassica oleracea*. *Nat Commun*. 2016b:7:13390. https://doi.org/10.1038/ncomms13390.

Golicz AA, Batley J, Edwards D. Towards plant pangenomics. *Plant Biotechnol J*. 2016a:14:1099–1105. https://doi.org/10.1111/pbi.12499.

Golicz AA, Bayer PE, Bhalla PL, Batley J, Edwards D. Pangenomics Comes of age: from bacteria to plant and animal applications. *Trends Genet.* 2020:36:132–145. https://doi.org/10.1016/j.tig.2019.11.006.

Gui S *et al.* A pan-Zea genome map for enhancing maize improvement. *Genome Biol.* 2022:23:178. https://doi.org/10.1186/s13059-022-02742-7.

Guo D *et al.* A pangenome reference of wild and cultivated rice. *Nature*. 2025a:642:662–671. https://doi.org/10.1038/s41586-025-08883-6.

Guo L *et al.* Super pangenome of *Vitis empowers* identification of downy mildew resistance genes for grapevine improvement. *Nat Genet.* 2025b:57:741–753. https://doi.org/10.1038/s41588-025-02111-7.

Guo N *et al.* A graph-based pan-genome of Brassica oleracea provides new insights into its domestication and morphotype diversification. *Plant Commun.* 2024:5:100791. https://doi.org/10.1016/j.xplc.2023.100791.

Guo W *et al.* A barley pan-transcriptome reveals layers of genotype-dependent transcriptional complexity. *Nat Genet.* 2025c:57:441–450. https://doi.org/10.1038/s41588-024-02069-y.

Guo W, Coulter M, Waugh R, Zhang R. The value of genotype-specific reference for transcriptome analyses in barley. *Life Sci Alliance*. 2022:5:e202101255. https://doi.org/10.26508/lsa.202101255.

He F *et al.* Pan-genomic analysis highlights genes associated with agronomic traits and enhances genomics-assisted breeding in alfalfa. *Nat Genet.* 2025:57:1262–1273. https://doi.org/10.1038/s41588-025-02164-8.

He Q *et al.* A graph-based genome and pan-genome variation of the model plant *Setaria*. *Nat Genet.* 2023:55:1232–1242. https://doi.org/10.1038/s41588-023-01423-w.

Hickey G *et al.* Genotyping structural variants in pangenome graphs using the vg toolkit. *Genome Biol*. 2020:21:35. https://doi.org/10.1186/s13059-020-1941-7.

Hickey G *et al.* Pangenome graph construction from genome alignments with Minigraph-Cactus. *Nat Biotechnol*. 2024:42:663–673. https://doi.org/10.1038/s41587-023-01793-w.

Hu W *et al.* Resequencing of 388 cassava accessions identifies valuable loci and selection for variation in heterozygosity. *Genome Biol.* 2021:22:316. https://doi.org/10.1186/s13059-021-02524-7.

Hu Z *et al.* EUPAN enables pan-genome studies of a large number of eukaryotic genomes. *Bioinformatics*. 2017:33:2408–2409. https://doi.org/10.1093/bioinformatics/btx170.

Huang Y *et al.* Pangenome analysis provides insight into the evolution of the orange subfamily and a key gene for citric acid accumulation in citrus fruits. *Nat Genet*. 2023:55:1964–1975. https://doi.org/10.1038/s41588-023-01516-6.

Huang Y *et al.* Multiscale pangenome graphs empower the genomic dissection of mixed-ploidy sugarcane species. *Science*. 2026:391:eadx1616. https://doi.org/10.1126/science.adx1616.

Hübner S *et al.* Sunflower pan-genome analysis shows that hybridization altered gene content and disease resistance. *Nat Plants.* 2019:5:54–62. https://doi.org/10.1038/s41477-018-0329-0.

Hufford MB *et al.* De novo assembly, annotation, and comparative analysis of 26 diverse maize genomes. *Science*. 2021:373:655–662. https://doi.org/10.1126/science.abg5289.

Ibañez VN, Quadrana L. Shaping inheritance: how distinct reproductive strategies influence DNA methylation memory in plants. *Curr Opin Genet Dev.* 2023:78:102018. https://doi.org/10.1016/j.gde.2022.102018.

Igolkina AA *et al.* A comparison of 27 *Arabidopsis thaliana* genomes and the path toward an unbiased characterization of genetic polymorphism. *Nat Genet*. 2025:57:2289–2301. https://doi.org/10.1038/s41588-025-02293-0.

Jayakodi M *et al.* Structural variation in the pangenome of wild and domesticated barley. *Nature*. 2024:636:654–662. https://doi.org/10.1038/s41586-024-08187-1.

Jayakodi M, Shim H, Mascher M. What are we learning from plant pangenomes? *Annu Rev Plant Biol.* 2025:76:663–686. https://doi.org/10.1146/annurev-arplant-090823-015358.

Jiao C *et al.* Pan-genome bridges wheat structural variations with habitat and breeding. *Nature*. 2025:637:384–393. https://doi.org/10.1038/s41586-024-08277-0.

Kang M *et al.* The pan-genome and local adaptation of *Arabidopsis thaliana*. *Nat Commun.* 2023:14:6259. https://doi.org/10.1038/s41467-023-42029-4.

Kawakatsu T *et al.* Epigenomic diversity in a global collection of *Arabidopsis thaliana* accessions. *Cell*. 2016:166:492–505. https://doi.org/10.1016/j.cell.2016.06.044.

Kaye AM, Wasserman WW. The genome atlas: navigating a new era of reference genomes. *Trends Genet.* 2021:37:807–818. https://doi.org/10.1016/j.tig.2020.12.002.

Khan AW *et al.* Super-pangenome by integrating the wild side of a species for accelerated crop improvement. *Trends Plant Sci.* 2020:25:148–158. https://doi.org/10.1016/j.tplants.2019.10.012.

Khan AW *et al.* Cicer super-pangenome provides insights into species evolution and agronomic trait loci for crop improvement in chickpea. *Nat Genet.* 2024:56:1225–1234. https://doi.org/10.1038/s41588-024-01760-4.

Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol.* 2019:37:907–915. https://doi.org/10.1038/s41587-019-0201-4.

Kopalli V, Arslan K, Morales-Díaz N, Zanini SF, Golicz AA. Toward a standardized framework for pangenome graph evaluation: assessing crop plant pangenome variation graph construction from multiple assemblies. *Gigascience*. 2025:14:giaf121. https://doi.org/10.1093/gigascience/giaf121.

Koren S *et al.* De novo assembly of haplotype-resolved genomes with trio binning. *Nat Biotechnol.* 2018:36:1174–1182. https://doi.org/10.1038/nbt.4277.

LaBarre BA *et al.* MethylToSNP: identifying SNPs in illumina DNA methylation array data. *Epigenetics Chromatin.* 2019:12:79. https://doi.org/10.1186/s13072-019-0321-6.

Lei L *et al.* Plant pan-genomics comes of age. *Annu Rev Plant Biol.* 2021:72:411–435. https://doi.org/10.1146/annurev-arplant-080720-105454.

Leventhal L, Ruffley M, Exposito-Alonso M. Planting genomes in the wild: *Arabidopsis* from genetics history to the ecology and evolutionary genomics era. *Annu Rev Plant Biol*. 2025:76:605–635. https://doi.org/10.1146/annurev-arplant-071123-095146.

Li H *et al.* Graph-based pan-genome reveals structural and sequence variations related to agronomic traits and domestication in cucumber. *Nat Commun.* 2022:13:682. https://doi.org/10.1038/s41467-022-28362-0.

Li H, Feng X, Chu C. The design and construction of reference pangenome graphs with minigraph. *Genome Biol*. 2020:21:265. https://doi.org/10.1186/s13059-020-02168-z.

Li J *et al.* Cotton pan-genome retrieves the lost sequences and genes during domestication and selection. *Genome Biol.* 2021:22:119. https://doi.org/10.1186/s13059-021-02351-w.

Li N, *et al.* Super-pangenome analyses highlight genomic diversity and structural variation across wild and cultivated tomato species. *Nat Genet*. 2023b:55:852–860. https://doi.org/10.1038/s41588-023-01340-y.

Li P, *et al.* Genomic insights into selection for heterozygous alleles and woody traits in Populus tomentosa. *Plant Biotechnol J*. 2023a:21:2002–2018. https://doi.org/10.1111/pbi.14108.

Li W, *et al.* Pan-genome analysis reveals the evolution and diversity of *Malus*. *Nat Genet*. 2025b:57:1274–1286. https://doi.org/10.1038/s41588-025-02166-6.

Li X *et al.* Large-scale gene expression alterations introduced by structural variation drive morphotype diversification in *Brassica oleracea*. *Nat Genet.* 2024:56:517–529. https://doi.org/10.1038/s41588-024-01655-4.

Li Y, *et al.* Panvariome and pangenome of 1,020 global peach accessions shed light on evolution patterns, hidden natural variations, and efficient gene discovery. *Mol Plant*. 2025a:18:995–1013. https://doi.org/10.1016/j.molp.2025.04.009.

Lian Q *et al.* A pan-genome of 69 Arabidopsis thaliana accessions reveals a conserved genome structure throughout the global species range. *Nat Genet.* 2024:56:982–991. https://doi.org/10.1038/s41588-024-01715-9.

Lian Q, Jiao W-B, Wang Y. Designing better crops with phased pangenomes. *Mol Plant.* 2025:18:1806–1808. https://doi.org/10.1016/j.molp.2025.08.014.

Liang Y-Y, *et al.* Pan-genome analysis reveals local adaptation to climate driven by introgression in Oak Species. *Mol Biol Evol*. 2025:42:msaf088. https://doi.org/10.1093/molbev/msaf088.

Liao W-W *et al.* A draft human pangenome reference. *Nature*. 2023:617:312–324. https://doi.org/10.1038/s41586-023-05896-x.

Lisch D. How important are transposons for plant evolution? *Nat Rev Genet.* 2013:14:49–61. https://doi.org/10.1038/nrg3374.

Liu M, *et al.* PPanG: a precision pangenome browser enabling nucleotide-level analysis of genomic variations in individual genomes and their graph-based pangenome. *BMC Geno*. 2024b:25:405. https://doi.org/10.1186/s12864-024-10302-5.

Liu T, Conesa A. Profiling the epigenome using long-read sequencing. *Nat Genet.* 2025:57:27–41. https://doi.org/10.1038/s41588-024-02038-5.

Liu Y *et al.* Pan-genome of wild and cultivated soybeans. *Cell*. 2020:182:162–176.e13. https://doi.org/10.1016/j.cell.2020.05.023.

Liu Y *et al.* SoyOmics: a deeply integrated database on soybean multi-omics. *Mol Plant.* 2023:16:794–797. https://doi.org/10.1016/j.molp.2023.03.011.

Liu Y, Siegmund KD, Laird PW, Berman BP. Bis-SNP: combined DNA methylation and SNP calling for Bisulfite-seq data. *Genome Biol.* 2012:13:R61. https://doi.org/10.1186/gb-2012-13-7-r61.

Liu Z, *et al.* Grapevine pangenome facilitates trait genetics and genomic breeding. *Nat Genet*. 2024a:56:2804–2814. https://doi.org/10.1038/s41588-024-01967-5.

Loegler V, Friedrich A, Schacherer J. Dynamics of genome evolution in the era of pangenome analysis. *Cell Genom*. 2026:6:101067. https://doi.org/10.1016/j.xgen.2025.101067.

Long W *et al.* Genome evolution and diversity of wild and cultivated rice species. *Nat Commun.* 2024:15:9994. https://doi.org/10.1038/s41467-024-54427-3.

Lynch RC *et al.* Domesticated cannabinoid synthases amid a wild mosaic cannabis pangenome. *Nature*. 2025:643:1001–1010. https://doi.org/10.1038/s41586-025-09065-0.

Ma W, *et al.* Gapless pangenome analyses reveal fast *Brassica rapa* subspeciation. *Science*. 2026:391:eady7590. https://doi.org/10.1126/science.ady7590.

MacPhillamy C, *et al.* DNA methylation analysis to differentiate reference, breed, and parent-of-origin effects in the bovine pangenome era. *GigaScience*. 2024:13:giae061. https://doi.org/10.1093/gigascience/giae061.

Manzano-Morales S, Liu Y, González-Bodí S, Huerta-Cepas J, Iranzo J. Comparison of gene clustering criteria reveals intrinsic uncertainty in pangenome analyses. *Genome Biol.* 2023:24:250. https://doi.org/10.1186/s13059-023-03089-3.

Marroni F, Pinosio S, Morgante M. Structural variation and genome complexity: is dispensable really dispensable? *Curr Opin Plant Biol.* 2014:18:31–36. https://doi.org/10.1016/j.pbi.2014.01.003.

Massonnet M, *et al.* Dissection of the Ren6 and Ren7 powdery mildew resistance loci in *Vitis piasezkii* DVIT2027 using phased parental–progeny genomes and intraspecific locus graph reconstruction. *G3 (Bethesda)*. 2025:15:jkaf250. https://doi.org/10.1093/g3journal/jkaf250.

Matthews CA, Watson-Haigh NS, Burton RA, Sheppard AE. A gentle introduction to pangenomics. *Brief Bioinform*. 2024:25:bbae588. https://doi.org/10.1093/bib/bbae588.

Meng Q *et al.* Pangenome analysis reveals yield- and fiber-related diversity and interspecific gene flow in *Gossypium barbadense* L. *Nat Commun.* 2025:16:4995. https://doi.org/10.1038/s41467-025-60254-x.

Mérot C, Oomen RA, Tigano A, Wellenreuther M. A roadmap for understanding the evolutionary significance of structural genomic variation. *Trends Ecol Evol.* 2020:35:561–572. https://doi.org/10.1016/j.tree.2020.03.002.

Miao Z, Yue J-X. Interactive visualization and interpretation of pangenome graphs by linear reference–based coordinate projection and annotation integration. *Genome Res.* 2025:35:296–310. https://doi.org/10.1101/gr.279461.124.

Michelmore RW, Meyers BC. Clusters of resistance genes in plants evolve by divergent selection and a birth-and-death process. *Genome Res.* 1998:8:1113–1130. https://doi.org/10.1101/gr.8.11.1113.

Mishra S, Srivastava AK, Khan AW, Tran L-SP, Nguyen HT. The era of panomics-driven gene discovery in plants. *Trends Plant Sci.* 2024:29:995–1005. https://doi.org/10.1016/j.tplants.2024.03.007.

Morales-Cruz A *et al.* Introgression among North American wild grapes (*Vitis*) fuels biotic and abiotic adaptation. *Genome Biol*. 2021:22:254. https://doi.org/10.1186/s13059-021-02467-z.

Mudge JM, Harrow J. The state of play in higher eukaryote gene annotation. *Nat Rev Genet.* 2016:17:758–772. https://doi.org/10.1038/nrg.2016.119.

Olsen KM, Wendel JF. A bountiful harvest: genomic insights into crop domestication phenotypes. *Annu Rev Plant Biol.* 2013a:64:47–70. https://doi.org/10.1146/annurev-arplant-050312-120048.

Olsen KM, Wendel JF. Crop plants as models for understanding plant adaptation and diversification. *Front Plant Sci.* 2013b:4:290. https://doi.org/10.3389/fpls.2013.00290.

Olson A *et al.* Gramene 2025: expanded comparative genomics and pathway resources, integrated search, and pan-genome portals for crop research. *Nucleic Acids Res*. 2026:54: D1720–D1732. https://doi.org/10.1093/nar/gkaf1260.

Paten B *et al.* Superbubbles, ultrabubbles, and Cacti. *J Comput Biol.* 2018:25:649–663. https://doi.org/10.1089/cmb.2017.0251.

Paten B, Novak AM, Eizenga JM, Garrison E. Genome graphs and the evolution of genome inference. *Genome Res*. 2017:27: 665–676. https://doi.org/10.1101/gr.214155.116.

Pedersen TL, Nookaew I, Wayne Ussery D, Månsson M. PanViz: interactive visualization of the structure of functionally annotated pangenomes. *Bioinformatics*. 2017:33:1081–1082. https://doi.org/10.1093/bioinformatics/btw761.

Pokrovac I, Pezer Ž. Recent advances and current challenges in population genomics of structural variation in animals and plants. *Front Genet*. 2022:13:1060898. https://doi.org/10.3389/fgene.2022.1060898.

Qiao Q, *et al.* Evolutionary history and pan-genome dynamics of strawberry (*Fragaria* spp.). *Proc Natl Acad Sci U S A.*. 2021:118: e2105431118. https://doi.org/10.1073/pnas.2105431118.

Qin P *et al.* Pan-genome analysis of 33 genetically diverse rice accessions reveals hidden genomic variations. *Cell*. 2021:184: 3542–3558.e16. https://doi.org/10.1016/j.cell.2021.04.046.

Rautiainen M, Marschall T. GraphAligner: rapid and versatile sequence-to-graph alignment. *Genome Biol.* 2020:21:253. https://doi.org/10.1186/s13059-020-02157-2.

Schreiber M, Jayakodi M, Stein N, Mascher M. Plant pangenomes for crop improvement, biodiversity and evolution. *Nat Rev Genet*. 2024:25:563–577. https://doi.org/10.1038/s41576-024-00691-4.

Secomandi S *et al.* Pangenome graphs and their applications in biodiversity genomics. *Nat Genet.* 2025:57:13–26. https://doi.org/10.1038/s41588-024-02029-6.

Shang L *et al.* A super pan-genomic landscape of rice. *Cell Res*. 2022:32:878–896. https://doi.org/10.1038/s41422-022-00685-z.

Sherman RM, Salzberg SL. Pan-genomics in the human genome era. *Nat Rev Genet*. 2020:21:243–254. https://doi.org/10.1038/s41576-020-0210-7.

Shi T *et al.* The super-pangenome of *Populus* unveils genomic facets for its adaptation and diversification in widespread forest trees. *Mol Plant.* 2024:17:725–746. https://doi.org/10.1016/j.molp.2024.03.009.

Shumate A, Salzberg SL. Liftoff: accurate mapping of gene annotations. *Bioinformatics*. 2021:37:1639–1643. https://doi.org/10.1093/bioinformatics/btaa1016.

Sibbesen JA *et al.* Haplotype-aware pantranscriptome analyses using spliced pangenome graphs. *Nat Methods*. 2023:20: 239–247. https://doi.org/10.1038/s41592-022-01731-9.

Sirén J, *et al.* Pangenomics enables genotyping of known structural variants in 5202 diverse genomes. *Science*. 2021:374: abg8871. https://doi.org/10.1126/science.abg8871.

Slabaugh E *et al.* Analysis of differential gene expression and alternative splicing is significantly influenced by choice of reference genome. *RNA*. 2019:25:669–684. https://doi.org/10.1261/rna.070227.118.

Soltis PS, Soltis DE. Plant genomes: markers of evolutionary history and drivers of evolutionary change. *Plants People Planet.* 2021:3:74–82. https://doi.org/10.1002/ppp3.10159.

Song J-M *et al.* Eight high-quality genomes reveal pan-genome architecture and ecotype differentiation of *Brassica napus*. *Nat Plants.* 2020:6:34–45. https://doi.org/10.1038/s41477-019-0577-7.

Stevenson KR, Coolon JD, Wittkopp PJ. Sources of bias in measures of allele-specific expression derived from RNA-seq data aligned to a single reference genome. *BMC Genomics*. 2013:14:536. https://doi.org/10.1186/1471-2164-14-536.

Sun C *et al.* RPAN: rice pan-genome browser for ∼3000 rice genomes. *Nucleic Acids Res*. 2017:45:597–605. https://doi.org/10.1093/nar/gkw958.

Sun X *et al.* Phased diploid genome assemblies and pan-genomes provide insights into the genetic history of apple domestication. *Nat Genet*. 2020:52:1423–1432. https://doi.org/10.1038/s41588-020-00723-9.

Tang D *et al.* Genome evolution and diversity of wild and cultivated potatoes. *Nature*. 2022:606:535–541. https://doi.org/10.1038/s41586-022-04822-x.

Tao Y *et al.* Extensive variation within the pan-genome of cultivated and wild sorghum. *Nat Plants.* 2021:7:766–773. https://doi.org/10.1038/s41477-021-00925-x.

Tao Y, Zhao X, Mace E, Henry R, Jordan D. Exploring and exploiting pan-genomics for crop improvement. *Mol Plant.* 2019:12: 156–169. https://doi.org/10.1016/j.molp.2018.12.016.

Teasdale LC *et al.* Pangenomic context reveals the extent of intraspecific plant NLR evolution. *Cell Host Microbe.* 2025:33:1291–-1305.e9. https://doi.org/10.1016/j.chom.2025.07.011.

Tettelin H *et al.* Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae* : implications for the microbial "pan-genome.". *Proc Natl Acad Sci U S A.* 2005:102:13950–-13955. https://doi.org/10.1073/pnas.0506758102.

Tian J *et al.* Teosinte ligule allele narrows plant architecture and enhances high-density maize yields. *Science*. 2019:365: 658–664. https://doi.org/10.1126/science.aax5482.

Van de Weyer A-L *et al.* A Species-wide inventory of NLR genes and alleles in *Arabidopsis thaliana*. *Cell*. 2019:178:1260–-1272.e14. https://doi.org/10.1016/j.cell.2019.07.038.

VanGessel CJ, *et al.* Ancient pangenomic origins of noncanonical NLR genes underlying the recent evolutionary rescue of a staple crop. *Sci Adv*. 2025:11:eady1667. https://doi.org/10.1126/sciadv.ady1667.

Voichek Y, Weigel D. Identifying genetic variants underlying phenotypic variation in plants without complete genomes. *Nat Genet.* 2020:52:534–540. https://doi.org/10.1038/s41588-020-0612-7.

Vondras AM *et al.* The genomic diversification of grapevine clones. *BMC Genomics*. 2019:20:972. https://doi.org/10.1186/s12864-019-6211-2.

Wang L *et al.* Pangenome analysis provides insights into legume evolution and breeding. *Nat Genet.* 2025:57:2052–2061. https://doi.org/10.1038/s41588-025-02280-5.

Wang T *et al.* Pan-genome analysis of 13 Malus accessions reveals structural and sequence variations associated with fruit traits. *Nat Commun.* 2023:14:7377. https://doi.org/10.1038/s41467-023-43270-7.

Wang W *et al.* Genomic variation in 3,010 diverse accessions of Asian cultivated rice. *Nature*. 2018:557:43–49. https://doi.org/10.1038/s41586-018-0063-9.

Wendel JF, Jackson SA, Meyers BC, Wing RA. Evolution of plant genome architecture. *Genome Biol.* 2016:17:37. https://doi.org/10.1186/s13059-016-0908-1.

Wong KHY *et al.* Towards a reference genome that captures global genetic diversity. *Nat Commun.* 2020:11:5482. https://doi.org/10.1038/s41467-020-19311-w.

Wu H, *et al.* Construction of the super pan-genome for the genus *Actinidia* reveals structural variations linked to phenotypic diversity. *Hortic Res*. 2025:12:uhaf067. https://doi.org/10.1093/hr/uhaf067.

Wu Z *et al.* Human pangenome analysis of sequences missing from the reference genome reveals their widespread evolutionary, phenotypic, and functional roles. *Nucleic Acids Res*. 2024:52:2212–2230. https://doi.org/10.1093/nar/gkae086.

Wulfridge P, Langmead B, Feinberg AP, Hansen KD. Analyzing whole genome bisulfite sequencing data from highly divergent genotypes. *Nucleic Acids Res*. 2019:47:e117. https://doi.org/10.1093/nar/gkz674.

Xia Z *et al.* Pan-genome and haplotype map of cassava cultivars and wild ancestors provide insights into its adaptive evolution and domestication. *Mol Plant.* 2025:18:1047–1071. https://doi.org/10.1016/j.molp.2025.05.014.

Yan H *et al.* Pangenomic analysis identifies structural variation associated with heat tolerance in pearl millet. *Nat Genet*. 2023:55:507–518. https://doi.org/10.1038/s41588-023-01302-4.

Yang T *et al.* Improved pea reference genome and pan-genome highlight genomic features and evolutionary characteristics. *Nat Genet*. 2022:54:1553–1563. https://doi.org/10.1038/s41588-022-01172-2.

Yao E, *et al.* GrainGenes: a data-rich repository for small grains genetics and genomics. *Database (Oxford)*. 2022:2022:baac034. https://doi.org/10.1093/database/baac034.

Ye X *et al.* Engineering the provitamin A (β-carotene) biosynthetic pathway into (carotenoid-free) rice endosperm. *Science*. 2000:287:303–305. https://doi.org/10.1126/science.287.5451.303.

Yildiz G *et al.* Graphical pangenomics-enabled characterization of structural variant impact on gene expression in *Brassica napus*. *Theor Appl Genet.* 2025:138:91. https://doi.org/10.1007/s00122-025-04867-2.

Yu X *et al.* Super pan-genome reveals extensive genomic variations associated with phenotypic divergence in *Actinidia*. *Mol Hortic.* 2025:5:4. https://doi.org/10.1186/s43897-024-00123-1.

Zhang H, *et al.* Super-pangenome analyses across 35 accessions of 23 Avena species highlight their complex evolutionary history and extensive genomic diversity. *Nat Genet*. 2025b:57:2276–2288. https://doi.org/10.1038/s41588-025-02294-z.

Zhang J, Zheng Y, Chen F. Phased telomere-to-telomere super-pangenome: definitive reference genome in plants. *Trends Plant Sci*. 2025a:31:266–269. https://doi.org/10.1016/j.tplants.2025.11.002.

Zhang L, *et al.* Solanaceae pan-genomes reveal extensive fractionation and functional innovation of duplicated genes. *Plant Commun*. 2025d:6:101231. https://doi.org/10.1016/j.xplc.2024.101231.

Zhang W, *et al.* methylGrapher: genome-graph-based processing of DNA methylation data from whole genome bisulfite sequencing. *Nucleic Acids Res*. 2025c:53:gkaf028. https://doi.org/10.1093/nar/gkaf028.

Zhang X *et al.* Haplotype-resolved genome assembly provides insights into evolutionary history of the tea plant *Camellia sinensis*. *Nat Genet.* 2021:53:1250–1259. https://doi.org/10.1038/s41588-021-00895-y.

Zhang Y, *et al.* Telomere-to-telomere *Citrullus* super-pangenome provides direction for watermelon breeding. *Nat Genet*. 2024b:56:1750–1761. https://doi.org/10.1038/s41588-024-01823-6.

Zhang Z, *et al.* A *k*-mer-based pangenome approach for cataloging seed-storage-protein genes in wheat to facilitate genotype-to-phenotype prediction and improvement of end-use quality. *Mol Plant*. 2024a:17:1038–1053. https://doi.org/10.1016/j.molp.2024.05.006.

Zhao K *et al.* Pangenome analysis reveals structural variation associated with seed size and weight traits in peanut. *Nat Genet.* 2025:57:1250–1261. https://doi.org/10.1038/s41588-025-02170-w.

Zhou Q, Wang Z, Li J, Sung W-K, Li G. MethHaplo: combining allele-specific DNA methylation and SNPs for haplotype region identification. *BMC Bioinform*. 2020:21:451. https://doi.org/10.1186/s12859-020-03798-7.

Zhou Y *et al.* The population genetics of structural variants in grapevine domestication. *Nat Plants.* 2019:5:965–979. https://doi.org/10.1038/s41477-019-0507-8.