



Accessing crop genetic diversity via pangenomics

Tessa R. MacNish^{1,2} · Venkataramana Kopalli³ · Silvia F. Zanini³ · Rod J. Snowdon⁴ · Agnieszka A. Golicz³ · David Edwards^{1,2}

Received: 17 June 2025 / Accepted: 27 February 2026 / Published online: 13 March 2026
© The Author(s) 2026

Abstract

With the increasing accuracy and decreasing cost of sequencing technology, the extent of structural variation (SV) and its importance in crop species has become increasingly evident. SVs such as insertions, deletions, and inversions have been associated with genetic variation of agronomically important traits and the diversification of crop species. Pangenomes aim to capture the genetic diversity of a species, population or genus, by incorporating the genomes of multiple individuals. The additional genetic diversity represented by a pangenome compared to a single-genome reference can aid the association of variation with traits and support crop improvement. Genus-wide pangenomes representing related crop species as well as their wild relatives can be used to identify and introduce novel genetic variation associated with agronomically important traits into crops. Pangenomes can aid crop improvement through pangenome assisted breeding (PAB) and genome editing. PAB is an adaption of marker assisted breeding that associates pangenome-based markers, including single nucleotide polymorphisms (SNPs) and SVs, with a trait of interest. Genome editing can use CRISPR/cas9 or similar tools to introduce or change the expression of agronomically important SVs. Pan-epigenomics is an emerging field that can complement pangenomics studies by identifying epigenetic modifications such as DNA methylation, histone modifications, and chromatin accessibility, which play important roles in regulating gene expression and have been shown to contribute to intraspecific diversity and agronomically important traits. We highlight the advances of crop pangenomics and their use in crop breeding and improvement.

Advances in DNA sequencing have revolutionised genomics

Our ability to read the genetic code of life has improved phenomenally over the last few decades, moving from manual radiolabelled Sanger-based DNA sequencing, through fluorescent labelled capillary based automated Sanger and then the introduction of massively parallel DNA sequencing (see

Batley and Edwards 2016 for a comprehensive review). The first eucaryotic genomes were assembled using automated Sanger sequencing (Kaul et al. 2000; Lander et al. 2001), requiring multimillion-dollar investment over several years. As technologies improved it became feasible to assemble multiple reference genomes for a species and differentiate between sequence/assembly errors and true biological variation (Bayer et al. 2017). These early studies demonstrated the previously suspected high level of presence absence variation in plant genomes (Li et al. 2014). As the cost of DNA sequencing continued to plummet and the accuracy of long read DNA sequencing improved, it became possible to examine genome variation on a species wide scale revealing the full extent of variation (Yao et al. 2015), the mechanisms leading to sequence duplication and loss as well as the impact that this variation has in both natural and domesticated population (see Zanini et al. 2022 for a literature review). Pangenomes are multi-individual genomic resources and can capture this genomic variation for more comprehensive analysis than single reference genomes permit. As our understanding of genomics continues to grow, and with it, our ability to edit genomes, the potential for

Communicated by Martin Mascher.

✉ David Edwards
dave.edwards@uwa.edu.au

¹ School of Biological Sciences, The University of Western Australia, Perth, WA 6009, Australia

² Centre for Applied Bioinformatics, The University of Western Australia, Perth, WA 6009, Australia

³ Department of Agrobiogenetics, IFZ Research Centre for Biosystems, Land Use and Nutrition, Justus Liebig University, Giessen, 35392 Giessen, Germany

⁴ Department of Plant Breeding, IFZ Research Centre for Biosystems, Land Use and Nutrition, Justus Liebig University, Giessen, 35392 Giessen, Germany

unprecedented scientific and societal advances becomes reality (see Tay Fernandez et al. 2022 for an in-depth review). Pangenomes will play an important part in this future as understanding the natural variation present in populations as well as the gene content of individuals to be edited are important for accurate and targeted genome modification.

Structural variation has significant impacts on crop evolution and adaptation—early evidence to today

Early comparative analyses of bacterial artificial chromosome (BAC) contigs (Morgante et al. 2005) and array-based, comparative genomic hybridization data in maize (Springer et al. 2009), and of short-read whole-genome resequencing datasets from soybean (Lam et al. 2010; McHale et al. 2012) indicated that gene copy number and presence absence variation were widespread and frequently associated to adaptive traits. While transposon activity is known to play a notable role in local SV occurrence in maize, giving rise to widespread tandem gene duplications (Springer et al. 2009), SV patterns observed among ancestral gene homologues in maize and soybean also indicated a role of ancient polyploidisation events in SV formation, reflecting similar observations in *Arabidopsis* (Blanc and Wolfe 2004). This suggested a direct role of polyploidisation as a driver of gene diversification in crop evolution, although empirical evidence is difficult to establish in ancient polyploids.

Although large structural inversions can also have a significant impact on recombination and are hence important for breeding efforts, the vast majority of SV in most crop plants involve presence-absence variants (PAV) caused by local transposon activity, small-scale chromosome rearrangements or by more substantial chromosome exchanges, particularly in polyploid species. One of the first major examples demonstrating the significant evolutionary impact of gene PAV caused by large-scale homoeologous chromosome exchanges in the course of polyploid crop evolution was discovered when the first genome assembly of the recent allopolyploid crop, *Brassica napus* (oilseed rape, canola, swede/rutabaga) was completed (Chalhoub et al. 2014). Widespread homoeologous chromosome exchanges and gene conversions were subsequently discovered in the de novo-assembled and resequenced genomes (Chalhoub et al. 2014; Samans et al. 2017; Hurgobin et al. 2017). Since *B. napus* is a recent crop species thought to have arisen in the past few thousand years, after few isolated allopolyploidisation events (Lu et al. 2019) and subsequent outcrossing with its diploid ancestors, the overall level of genetic diversity in the species is expected to be limited. Limited diversity should inhibit evolutionary or breeding success. Paradoxically however, *B. napus* has become one of the most globally successful oilseed crop

species and achieved broad ecotype diversification and eco-geographical adaptation to diverse climatic zones on all continents. A better understanding of the extent, mechanisms and consequences of de novo diversification through homoeologous genome interactions and rearrangements in *B. napus* gave rise to the hypothesis that widespread SV formation during and after polyploidisation is a driver of de novo crop diversity following the strong bottlenecks of polyploidisation and domestication, and that existing and ongoing SV formation enabled historical crop evolution and could be a key to future adaptation (Samans et al. 2017). Chawla et al. (2020) showed that SV alone can explain the diversification of *B. napus* into differentiated ecotypes and crop forms, while other studies linked specific intragenic deletions/insertions with disease resistance or flowering behaviour (Gabur et al. 2020; Vollrath et al. 2021a, 2021b). Later assemblies confirmed that the enlarged tuber of *B. napus* swede forms and the related tuberous morphotypes in the related allopolyploid *B. juncea* likely arose by direct human selection from the product of homoeologous chromosome exchanges that led to CNV among the genes responsible for this trait. Interestingly, SVs in *B. napus* have been shown to arise at a much higher rate than expected, with some genotypes having new variants arising in each generation (Orantes-Bonilla et al. 2022).

Beyond well-studied examples such as *B. napus*, the growing availability of high-quality genome and pangenome sequence assemblies for crop species is providing evidence to support a major role of genome-wide SV in crop evolution and diversification. An investigation of genome-wide gene PAV using exon capture across a panel of 434 wheat accessions revealed the presence of widespread chromosome introgressions, and gene loss or duplication caused by non-reciprocal homoeologous exchanges (Heuberger et al. 2024), reminiscent of the patterns reported in *B. napus*, and with related associations to ecogeographical diversification into groups with divergent phenological adaptation. Large chromosome introgressions were found to be frequent and to impact traits relevant for wheat adaptation and breeding, including important disease resistance traits. In many cases, introgressions originated from wild grass species. As in *B. napus*, evidence was also found that SVs affecting phenological genes have facilitated both global and local adaptation to ecogeographical regions (Jiao et al. 2025). Similarly extensive occurrence of gene presence absence variation as a consequence of homoeologous exchanges have also been reported in the hexaploid oat genome (Peng et al. 2022). Another example of this phenomenon was recently reported in natural populations of white clover (*Trifolium repens*), a globally important allopolyploid that has widespread natural populations and is also widely cultivated as a grassland crop. In an analysis of more than 2600 *T. repens* plants collected from six continents, (Battlay et al. 2024) revealed that large

structural variants had imparted strong signatures of parallel adaptation to climatic zones.

Interestingly, new data from ancient polyploidisation examples indicate that polyploidisation is a strong evolutionary force under changing abiotic and biotic stress conditions due to its role in generating diversity (see van der Peer et al. 2021 for a comprehensive review). For example, whole-genome resequencing of 363 diverse accessions of sorghum, a major tropical cereal crop with a relatively small and simple diploid genome, showed that evolution into divergent crop use types and the emergence of variants for adaptation traits such as day-length are attributable to SV, some of which date back to ancient whole-genome duplication events (Zhang et al. 2024c). Overall, SVs were shown to have a disproportionate impact on trait variation, supporting a very early and ongoing association of SV with the emergence and selection of traits and thus potentially a major role in crop evolution. These results support work by Songsomboon et al. (2021), who observed cluster-specific deletions of genes related to biotic and abiotic stress responses and suggested a role of such deletions in local adaptation.

It is known that SV can readily occur through aberrant DNA repair after non-homologous end-joining, which requires only low- to medium-level sequence similarity at breakpoints (see the review by Saxena et al. 2014) and thus acts not only across interacting homoeologous chromosome segments but also in relation to tandem repeats. This may be an explanation for why abundant classes of resistance genes that are found in tandemly repeated clusters are frequently impacted by CNV, and why local tandem gene duplications are commonly observed in large crop genomes that have a high density of repetitive elements (Morgante et al. 2005; Jayakodi et al. 2023). Associations of SV impacting resistance gene clusters and corresponding disease traits are therefore not unexpected and have been reported in many important crops including soybean (McHale et al. 2012), sorghum (Zheng et al. 2011), rice (Fuentes et al. 2019) and tomato (see Andolfo et al. 2021 for a review).

The body of evidence supporting the impact of SV on crop evolution and adaptation continues to grow as pangenomic datasets and related trait observations are associated with these variants. The continuing discovery of trait-relevant SV and their implementation for quantitative genetic analyses and predictive breeding is likely to further increase our knowledge of their impact and utility throughout the coming decade.

The development of pangenome methods

A pangenome captures the genetic diversity observed across a group of related individuals, aiming to be broadly representative of the wider population rather than exhaustive.

This definition reflects that no pangenome can include every allele present in a species, but well-designed datasets provide a close approximation of population level diversity. The concept was first introduced in *Streptococcus agalactiae* (Tettelin et al. 2005), classifying genes as core (present in all individuals) and dispensable (absent in at least one).

Pangenomes capture complex genomic polymorphisms including structural variations (SVs), PAVs, and copy number variations (CNVs), key determinants of phenotypic variation across crop species. However, these resources often do not represent the full diversity within a species due to limitations in sampling, sequencing depth, or the difficulty of assembling repetitive sequences (see the reviews Sarashetti et al. 2024; Secomandi et al. 2025; Tørresen et al. 2019 for more details). Earlier studies focused mainly on gene level presence/absence matrices, whereas most current approaches use chromosome level assemblies that can represent large SVs and both genic and intergenic diversity. Chromosome scale pangenomes have now been constructed for several major crop species such as rice, maize, wheat, soybean, and rapeseed (Hirsch et al. 2014; Li et al. 2014; Montenegro et al. 2017; Song et al. 2020; Wang et al. 2018).

The following sections describe major methodological paradigms for pangenome construction, broadly categorized into linear and graph-based frameworks, as well as other approaches and are summarized in Table 1.

Linear pangenome approaches

Linear pangenomes represent genetic diversity in a linear coordinate framework, extending or comparing assemblies to identify shared and variable regions. These methods are straightforward to implement and require relatively little computational power, but they cannot fully represent alternative haplotypes or large-scale structural rearrangements. Linear pangenomes can be generated either by an iterative mapping and assembly, or by comparison of independently assembled genomes; both differ from graph models, which encode multiple paths within a single coordinate space.

Iterative mapping and assembly

The Iterative Mapping and Assembly strategy constructs a pangenome by repeatedly aligning reads from multiple individuals to a single reference, assembling unmapped reads, and integrating these sequences into the reference. This approach allows the reference to grow as more individuals are added, progressively revealing non-reference genomic content. This strategy was implemented by Golicz et al. (2016) to generate the first *Brassica oleracea* pangenome, employing Bowtie2 (Langmead and Salzberg 2012) for read alignment, MaSuRCA (Zimin et al. 2013) for genome assembly and analysis, and SGSGeneLoss

Table 1 Pangenome construction methods in both linear and graph-based pangenomes, highlighting their workflow, strengths, and limitations. Selected supporting studies illustrate their applications in plant research

Method	Main steps	Strengths	Limitations	First application in plants
Linear Pangenome	Iterative Mapping & Assembly	Assemble genomes from multiple accessions → Map to reference → Call variants → Add novel sequences as unassembled contigs to the improved reference	Computationally efficient, good for small-scale studies	Limited detection of large SVs, reference-bias, Lack of contiguity /sequence context
De novo Assembly	Assemble genomes independently → Compare whole genomes → Call variants → Merge structural variations	Captures full genome diversity, unbiased by reference	Computationally intensive, requires high-quality long reads	(Y. Li et al. 2014) <i>Glycine soja</i> (Wild Soybean)
Graph Pangenome	Reference-based genetic variation graphs Align assemblies to a reference → Call variants → Construct graph from the variants	Efficient for genotyping and variant discovery, captures known genetic variation	Limited to known variants, may miss novel variations	(Liu et al. 2020) <i>Glycine max</i> (Soybean) Tool: vg toolkit (Garrison et al. 2018)
Reference-Based Pangenome	Align assemblies to a reference → Call variants → Construct variation graph from the variants	Preserves reference context, efficient for large datasets	Reference bias, may miss novel sequences	(Shang et al. 2022) <i>Oryza Sativa</i> (Rice) Tool: Minigraph (H. Li et al. 2020)
Reference-Unbiased Pangenome	Perform all-vs-all alignment of assemblies → Call variants → Construct graph without reference	Captures all structural variations, suitable for diverse species	High computational cost, complex to interpret	No published plant study yet
Reference-based Practical Haplotype Graph (PHG)	Define reference ranges → Upload haplotypes from assemblies → Call variants → Construct graph database	Efficient data storage, enables imputation for low-coverage sequences	Represents haplotypes instead of full sequences, visualization challenges	(Jensen et al. 2020) <i>Sorghum</i>

(Golicz et al. 2015) for gene PAV detection. Because outputs remain linear FASTA and VCF files anchored to reference coordinates, these workflows are directly compatible with conventional read mapping, variant calling, and gene annotation pipelines. The approach is computationally efficient and compatible with standard pipelines, but remains reference anchored and cannot capture large SVs or multiple haplotypes at a locus.

De Novo assembly based pangenomes

Methods based on full de novo assemblies, independently assemble each genome and compare them to delineate core and dispensable regions, avoiding reference bias and enabling discovery of novel loci. Early implementations relied on short read assemblies that were fragmented and structurally incomplete. With the advent of long read and HiFi sequencing, as well as scaffolding methods such as Hi-C, optical mapping, and haplotype phasing, de novo assemblies became chromosome level and highly contiguous. This technological shift transformed de novo assembly-based pangenomes from fragmented collections of contigs into accurate, reference free genomic resources compatible with downstream graph-based analyses.

The first plant pangenome based on de novo assemblies, built for *Glycine soja* (Li et al. 2014), used short reads from seven individuals but was limited in structural resolution. This early example illustrated how short read data restricted detection of long insertions, tandem duplications, and other complex SVs. Building on this foundation, subsequent pangenome studies, for example the potato pangenome (Tang et al. 2022) leveraged HiFi sequencing and Hifiasm (Cheng et al. 2021) assemblies to construct high-quality genomes for comparative analysis. Each genome was annotated prior to pangenome construction using MAKER2 (Holt and Yandell 2011), after which gene models were compared across accessions to define core and variable loci. Subsequent analyses focused on disease-resistance genes, which were re-annotated and classified using nucleotide-binding repeat (NLR) gene annotations (Steuernagel et al. 2020).

Graph based pangenome approaches

Graph based pangenomes represent genomic sequences and their variations in a non-linear structure. Nodes correspond to sequence segments, edges connect adjacent regions, and paths trace genome specific traversal (Zanini et al. 2022). These graphs encode sequence and structural variation in one unified framework, allowing direct mapping between linear and graph coordinates.

Reference based graphs

Reference based graphs anchor variation to an existing reference genome and integrate new sequences via alignment and variant calling.

Reference based genetic variation graph approach applied in tomato pangenome project by Zhou et al. (2022) combined Minimap2 (Li 2018), a fast long-read aligner for mapping sequences to large genomes, for alignment, DeepVariant (Poplin et al. 2018), a neural network-based variant caller that infers SNPs and indels from read alignments, and multiple SV callers (Sniffles, SVIM, CuteSV) merged with SURVIVOR (Jeffares et al. 2017), a toolkit for merging and comparing structural-variant call sets. Graphs were built using the vg toolkit (Garrison et al. 2018), a software suite for constructing, manipulating, and mapping to variation graphs. The subsequent graphs can be used to genotype SVs in larger population sequenced with short reads. The use of high-quality long-read assemblies for the backbone better represents the complex genomic variation found in plants, while the use of short reads for genotyping reduces sequencing costs. However, as a reference-based method, genetic variation graphs are biased to previously detected variants and may overlook novel or complex SVs.

Minigraph (Li et al. 2020), extends this concept to hundreds of assemblies, aligning each to a reference to automatically detect insertions, deletions, and rearrangements. Graphs made using Minigraph can be scaled to large datasets such as in the reference-based pangenome graphs constructed from 251 rice assemblies (Shang et al. 2022). This method preserves reference coordinate systems and scales well with many genomes, but the drawback is that it remains biased toward the reference and may miss diverged or novel sequences not present in the anchor genome (Kopalli et al. 2025, accepted). Minigraph-Cactus (Hickey et al. 2024), which integrates the Minigraph and Cactus frameworks, first uses Minigraph to generate a guide tree that defines the order of genome alignment and then applies the Cactus multiple-genome aligner to build a hierarchical graph based on these alignments. This combination enables a more accurate incorporation of new sequences, retaining a larger proportion of non-reference sequence while preserving the reference coordinate system and improving representation of complex rearrangements. Although its application in plants is still emerging, a recent study in rice used Minigraph-Cactus to successfully resolve multiallelic tandem LTR-retrotransposon, demonstrating its utility for complex repetitive loci (Morales-Díaz et al. 2025).

Reference unbiased graphs

Reference unbiased approaches treat all genomes equally and avoid anchoring variation to a single reference. The

PanGenome Graph Builder (PGGB) (Garrison et al. 2024), constructs reference-free pangenomes through all-vs-all sequence alignments and graph induction, capturing a broad spectrum of structural variation without introducing reference bias. While successfully applied to human and bacterial datasets, its application to plant genomes is computationally intensive due to abundance of repetitive elements. To mitigate this, PGGB can operate at the chromosome level, reducing runtime and memory demands. This strategy enables detailed representation of complex rearrangements, although inter-chromosomal variation may be missed and the resulting graphs can be difficult to interpret because of their high connectivity. Benchmarking across plant datasets (Kopalli et al. 2025, accepted) showed that PGGB generates highly resolved graphs that capture extensive structural diversity, but at substantial computational cost. Benchmarking across plant datasets (Kopalli et al. 2025, accepted) showed that PGGB can capture extensive structural diversity, but at substantial computational cost. However, unlike Minigraph and Minigraph-Cactus, which maintain distinct paths for duplicated or repetitive segments, PGGB often merges copy-number variable regions such as segmental duplications and VNTRs into a single shared path traversed by multiple haplotypes, simplifying graph topology but reducing explicit repeat resolution (Liao et al. 2023).

Comparative performance of pangenome graph construction methods

Studies in humans and cattle (Andreace et al. 2023; Leonard et al. 2023) have compared pangenome graph construction methods including Minigraph, Minigraph-Cactus, and PGGB. These studies found that Minigraph uses significantly less CPU, memory, and time than Minigraph-Cactus and PGGB, but does not capture small variants such as SNPs or short insertions and deletions and is affected by reference bias. Consequently, Minigraph-Cactus and PGGB retained up to five times more sequence from non-reference accessions (Leonard et al. 2023). PGGB constructed the most complex graphs, identifying the largest number of SNPs and haplotype paths, followed by Minigraph-Cactus and Minigraph (Andreace et al. 2023). PGGB also showed the most stable results across repeated runs, whereas Minigraph-Cactus graphs varied slightly and Minigraph graphs depended on the order in which accessions were added. While PGGB performs all-vs-all comparisons and removes reference bias, the entire graph must be rebuilt to incorporate new genomes, unlike Minigraph and Minigraph-Cactus, which can incrementally add assemblies.

Comparable benchmarking in plants has shown consistent trends. In this study by Kopalli et al. 2025, (accepted), systematic comparisons of Minigraph, Minigraph-Cactus, and PGGB using simulated and real-world assemblies in

Sorghum bicolor demonstrated that pairwise callers such as SyRI (Goel et al. 2019) and SVIM-asm (Heller and Vingron 2020) achieved slightly higher precision in controlled simulations, but graph-based methods captured a broader range of structural variants, including complex and multiallelic events, in real-world data. Graph-based pangenomes built with Minigraph-Cactus and PGGB, also showed improved read-mapping accuracy compared with linear references. However, PGGB required extremely high computational resources; for example, building mapping indexes for a graph of seven *Sorghum bicolor* genomes needed over 2 TB of memory. The same study also extended benchmarking to soybean, barley and oilseed rape, where trends largely mirrored those observed in *Sorghum bicolor*.

Together, these findings demonstrate that the most appropriate pangenome graph construction method depends on the computational resources available and the intended downstream analysis.

Pangenomes have transformed population-scale genomics, improving variant detection, read alignment, and expression quantification (Yildiz et al. 2025). Current benchmarking in crops (Kopalli et al. 2025, accepted) indicates that while graph pangenome approaches outperform linear models in accuracy, challenges remain in scalability, computational efficiency, and graph interpretation. Emerging solutions such as rpvg (Sibbesen et al. 2023) show promise for transcript and population scale analyses.

Other pangenome building approaches

The Practical Haplotype Graph (PHG) is a graph-based framework designed to represent genetic diversity using shared haplotypes instead of complete nucleotide-level assemblies. It captures variation as a network of consensus haplotype blocks that summarize sequence diversity across individuals, allowing efficient genotype imputation and genomic prediction. Developed for crop breeding (Jensen et al. 2020), the reference-based PHG method used in a sorghum pangenome study involved aligning sequence data to the sorghum BTx623 reference genome using BWA MEM (Li 2013). Variants were called with the HaplotypeCaller pipeline (Sentieon DNaseq, 2018), and the results used to construct consensus haplotypes. These haplotypes were clustered and condensed to form a compact set of consensus sequences. The pangenome graph is built by mapping these haplotypes using a hidden Markov model (HMM) and Viterbi algorithm (Rabiner 1989) and is stored in a relational database, allowing for efficient analysis and imputation of low-coverage genomic data. The PHG approach supports tasks such as genotyping and genomic prediction, helping to organize and analyze large genomic datasets (<https://bitbucket.org/bucklerlab/practicalhaplotypegraph/wiki/Home>).

PanTools (Sheikhizadeh et al. 2018) represents genomes using a generalized De Bruijn graph. It combines pangenome representation with functional and comparative genomics, enabling the storage and annotation of multiple genomes, orthology inference, and downstream analyses such as gene family clustering, functional annotation propagation, and syntenic block comparison.

Visualization tools for pangenome graphs

Visualization tools enable interpretation of complex pangenome graph structures. As pangenome graphs continue to increase in size and complexity, effective visualization is essential for exploring structural variants, inspecting graph topology, and validating genome alignments. These tools allow users to translate abstract graph data into interpretable visual formats, helping to identify genomic regions that differ between individuals and to assess connectivity within the graph.

ODGI (Optimized Dynamic Genome/Graph Implementation) (Guarracino et al. 2022) provides a comprehensive toolkit for graph manipulation and statistical analysis within GFA-based pangenomes. It supports editing, sorting, and dynamic updates of graphs, as well as interactive visualizations that facilitate the identification of genetic variation and haplotype structures. ODGI can also be used to edit graphs such as PGGB and Minigraph-Cactus by removing complex regions, ordering nodes, and merging segments; however, effective use of these features requires detailed knowledge of the tool's functions, parameters, and the overall structure of the pangenome graph (Guarracino et al. 2022).

Panache, a web-based platform, provides an intuitive interface for visualizing pangenomes, enabling users to explore gene content, structural variants, and genome alignments (Durant et al. 2021). Panache requires pangenome graphs to be linearized prior to visualization, where each continuous string of sequence is represented as a linear block. By linearizing the graph pangenome, some SV information is lost; however, the linear pangenome gains a stable coordinate system, for reproducible navigation, and the graph pangenome can be more easily displayed and interpreted.

Other visualization tools have been developed to address specific challenges associated with pangenome graph interpretation. Bandage (Wick et al. 2015) provides an accessible graphical interface for viewing small to medium-sized GFA graphs, offering an overview of graph topology and connectivity, though it becomes less effective with large, complex plant genomes. More recently, PanGraphViewer (Yuan et al. 2023) introduced a dedicated interface for pangenome graphs, supporting GFA and GFAv2 formats, integration with annotations, and comparison of multiple individuals. Together, these visualization platforms provide

complementary solutions that balance scalability, interactivity, and usability for exploring graph-based pangenomes.

Future directions in pangenome construction

Taken together, the comparisons above show that the choice between linear, graph-based, and haplotype-based frameworks is driven by study goals, data quality, and available computational resources. Each approach offers complementary strengths: linear methods provide simplicity and compatibility with existing tools, graph-based methods capture complex structural variation and enable improved mapping, and haplotype-based frameworks such as PHG allow efficient imputation for breeding applications. However, all current strategies face challenges in scalability, standardization, and interoperability.

Progress in pangenomics therefore depends on improving the efficiency of assembly and graph construction, enabling incremental updates as new genomes become available, and developing visualization tools that are more user-friendly and capable of preserving structural and functional information. In parallel, consistent formats and annotation practices are needed so that gene models, SVs, and presence/absence information can be exchanged and compared across methods. Integrating genomic, transcriptomic, and epigenomic data within unified pangenome frameworks and leveraging haplotype-based approaches for downstream applications will help move plant pangenomics toward comprehensive, application-ready resources for biological discovery and crop improvement.

Current state of pangenomics in crops

Graph-based pangenomes are considered the gold standard for pangenome construction because they can represent genome-wide haplotypes, store reference genome information alongside the variation between individuals, and represent multiple species, including wild relatives of domesticated crops (see Wang et al. 2023 for a review on the topic; Wang et al. 2024). Due to these benefits, an increasing number of graph-based pangenomes have been constructed for crop species including wheat, barley, tomato, and kiwifruit (Bayer et al. 2022a; Jayakodi et al. 2024; Li et al. 2023a; Wang et al. 2024). Bayer and colleagues (2022a) constructed a graph pangenome representing 16 bread wheat accessions and were able to identify the sequences that were conserved across all accessions, present only in some accessions, and unique to one accession. A graph pangenome representing 14 kiwifruit accessions found that 53% of gene families were variable between individuals (Wang et al. 2024). Within these SVs, the authors identified a variant in the promoter of *AcBCM*, associated with fruit color and increased chlorophyll

accumulation. A graph pangenome in maize (Gui et al. 2022) found that 37.36% of SVs were in low linkage disequilibrium with SNPs indicating that these SVs represented genetic variation that could not be identified using SNPs alone. Additionally, these SVs had higher heritability scores and explained 14% more phenotypic variation than SNPs and were therefore more likely to be functionally important. A trait association analysis using the same maize graph pangenome (Gui et al. 2022) supported this conclusion, with 47.5% of gene PAVs significantly associated with an agronomically important trait. The additional heritability and the significant genomic regions associated with these SVs could not have been identified using a single genome reference or by the de novo pangenome constructed in the same study, demonstrating the importance of using graph pangenomes alongside single-genome and de novo pangenome references to identify all agronomically important genomic variation.

Genus-wide pangenomes aid crop improvement by identifying novel sources of genetic variation that can be introduced into crop species through hybridization and genome editing (see Khan et al. 2020 for a comprehensive review). Modern crop varieties have declined in genetic diversity (see Khoury et al. 2022 for a review on the topic), making these genus-wide pangenomes an important resource for crop improvement. In recent years, multi-species and genus-wide pangenomes have been constructed for major crops as well as for some minor crops (Table 2). A rice pangenome representing 16 domesticated and wild *Oryza* species identified an additional 63,881 gene families present in the wild relatives (Long et al. 2024). The pangenome consisted of 57% dispensable gene families, 33% species-specific gene families, and 10% core gene families. Disease resistance genes were found to be less diverse in the cultivated rice accessions than in their wild relatives, demonstrating that the wild rice accessions could be a novel source of disease resistance. A genus-wide pangenome of *Glycine* found that, after duplicated genes were removed, approximately 70% of the genes identified in the wild perennial relatives were not present in annual domesticated soybean (Zhuang et al. 2022). This *Glycine* pangenome was constructed with 26 accessions consisting of 8 domesticated and wild species and demonstrated novel genetic variation within wild relatives. Su and colleagues (2024) constructed an apple pangenome including cultivated and wild varieties. A total of 63% of gene family sets were dispensable, and the wild species gene annotations were enriched for traits including oxidative phosphorylation, pentose metabolic process, and response to salt. The cultivated apples had a greater number of disease resistance genes due to gene retention following repeat hybridization/introgression with multiple wild relatives. These studies demonstrate the ability of genus-wide pangenomes to identify novel genetic variation that can improve

agronomically important traits and at the same time improve genetic diversity in crops through hybridization.

The quality of alignments to graph and genus-wide pangenomes depends on the quality and relatedness of sequences, which can lead to short, low quality or highly divergent sequences being absent from the final pangenome. Minigraph (Li et al. 2020) requires strong colinear chains for alignment and can fail to align short or highly divergent sequences. A *Brassica* genus-wide pangenome constructed using minigraph (MacNish et al. 2025) found that *B. napus* and *B. oleracea* shared more sequence than the more closely related *B. napus* and *B. rapa* due to the lower quality of the *B. rapa* genomes used. In addition, the accession with the highest number of unique sequences was the reference genome ZS11, which could indicate reference bias. All other reference-based graph construction methods such as Minigraph-Cactus (Hickey et al. 2024) and PHG (Bradbury et al. 2022) risk failing to align the highly divergent sequences found in genus-wide pangenomes, due to reference bias. Genus-wide pangenomes constructed through reference-based iterative assembly (He et al. 2021; van Workum et al. 2024) share this limitation. Reference unbiased methods such as de novo pangenome construction may be better suited to identifying novel or highly divergent sequences. Gui and colleagues (2022) constructed a genus-wide maize graph pangenome which had a mapping rate of 99.4% and alignment identities of 91.1% for individuals not included in the pangenome, indicating that the graph pangenome represented the majority of genetic variation within the population. Therefore, genus-wide and graph pangenomes can represent the genetic repertoire of a population depending on the species and construction methods used. However, many pangenomes studies do not report the percent of sequence that is not represented by the pangenome, which makes it challenging to know which pangenome resources accurately represent the whole population and which methods are best for each species or genus.

The application of pangenomes for crop improvement

The SVs identified in pangenomes contribute to gene expression and trait variation (Jeffares et al. 2017; Yuanyuan Zhang et al. 2024a) and have been significantly associated with agronomically important traits that can be used for crop improvement (Alonge et al. 2020). Pangenome assisted breeding (PAB) is an adaption of marker assisted breeding that associates pangenome-based markers, including SNPs and SVs, with a trait of interest. Pangenome-based markers can be identified by assembling a pangenome, mapping reads to the pangenome and calling SNPs and SVs (Chen et al. 2023a; Chen et al. 2023). Once pangenome markers

Table 2 Representative multi-species and genus-wide pangenome studies in crops and whether these pangenomes included accessions representing the wild relatives of the domesticated crop species

Genus	Species	Crop	Number of accessions	Pangenome type	Wild accessions	References
<i>Brassica</i>	<i>B. rapa</i> , <i>B. nigra</i> , <i>B. oleracea</i> , <i>B. napus</i> , <i>B. juncea</i> , and <i>B. carinata</i>		15	Iterative mapping and assembly	No	Z. He et al. 2021
<i>Brassica</i>	<i>B. napus</i> , <i>B. oleracea</i> , and <i>B. rapa</i>		41	Graph	No	MacNish et al. 2025
<i>Camellia</i>	<i>C. sinensis</i> and <i>C. assamica</i>	Tea	11	Gene-based	Yes	Tariq et al. 2024
<i>Cicer</i>	<i>C. arietinum</i> , <i>C. reticulatum</i> , <i>C. echinospermum</i> , <i>C. bijugum</i> , <i>C. judaicum</i> , <i>C. pinnatifidum</i> , <i>C. yamashitae</i> , <i>C. chorassanicum</i> and <i>C. cuneatum</i>	Chickpea	10	Graph	Yes	Khan et al. 2024
<i>Citrullus</i>	<i>C. naudinianus</i> , <i>C. colocynthis</i> , <i>C. rehmi</i> , <i>C. ecirrhosus</i> , <i>C. amarus</i> , <i>C. mucospermus</i> , and <i>C. lanatus</i>	Watermelon	28	Graph	Yes	Yilin Zhang et al. 2024b
<i>Citrullus</i>	<i>C. lanatus</i> , <i>C. mucospermus</i> , <i>C. amarus</i> , and <i>C. colocynthis</i>	Watermelon	547	Gene-based	Yes	Wu et al. 2023
<i>Cucumis</i>	<i>C. sativus</i> and <i>C. sativus</i>	Cucumber	12	Graph	Yes	H. Li et al. 2022
<i>Glycine</i>	<i>G. max</i> and <i>G. soja</i>	Soybean	1110	Iterative mapping and assembly	Yes	Bayer et al. 2022b
<i>Glycine</i>	<i>G. max</i> , <i>G. soja</i> , <i>G. falcata</i> , <i>G. stenophita</i> , <i>G. cyrtoloba</i> , <i>G. syndetika</i> , <i>G. tomentella</i> , and <i>G. dolichocarpa</i>	Soybean	26	Gene-based	Yes	Zhuang et al. 2022
<i>Hordeum</i>	<i>H. bulbosum</i> and <i>H. vulgare</i>	Barley	33	Graph	Yes	Feng et al. 2025
<i>Hordeum</i>	<i>H. vulgare</i> and <i>H. spontaneum</i>	Barley	76	Graph	Yes	Jayakodi et al. 2024
<i>Lactuca</i>	<i>L. sativa</i> , <i>L. serriola</i> , <i>L. saligna</i> , and <i>L. virosa</i>	Lettuce	474	Iterative mapping and assembly	Yes	van Workum et al. 2024
<i>Malus</i>	<i>M. domestica</i> , <i>M. baccata</i> , <i>M. sieversii</i> , and <i>M. sylvestris</i>	Apple	12	Gene-based	Yes	Su et al. 2024
<i>Musa and Ensete</i>	<i>M. acuminata</i> , <i>E. ventricosum</i> , <i>M. itinerans</i> , <i>M. balbisiana</i> , <i>Musa x paradisiaca</i> , and <i>M. textilis</i>	Banana	15	Iterative mapping and assembly	No	Rijzaani et al. 2022
<i>Oryza</i>	16 species	Rice	17	Gene-based	Yes	Long et al. 2024
<i>Panicum</i>	<i>P. miliaceum</i>	Broomcorn millet	32	Graph	Yes	J. Chen et al. 2023a
<i>Pyrus</i>	<i>P. bretschneideri</i> , <i>P. sinkiangensis</i> , <i>P. communis</i> , <i>P. bretschneideri</i> , <i>P. betulifolia</i> , <i>P. pyrifolia</i> , and <i>P. ussuriensis x communis</i>	Pear	7	de novo	Yes	Ding et al. 2024
<i>Solanum</i>	60 species	Potato	296	de novo	Yes	Bozan et al. 2023

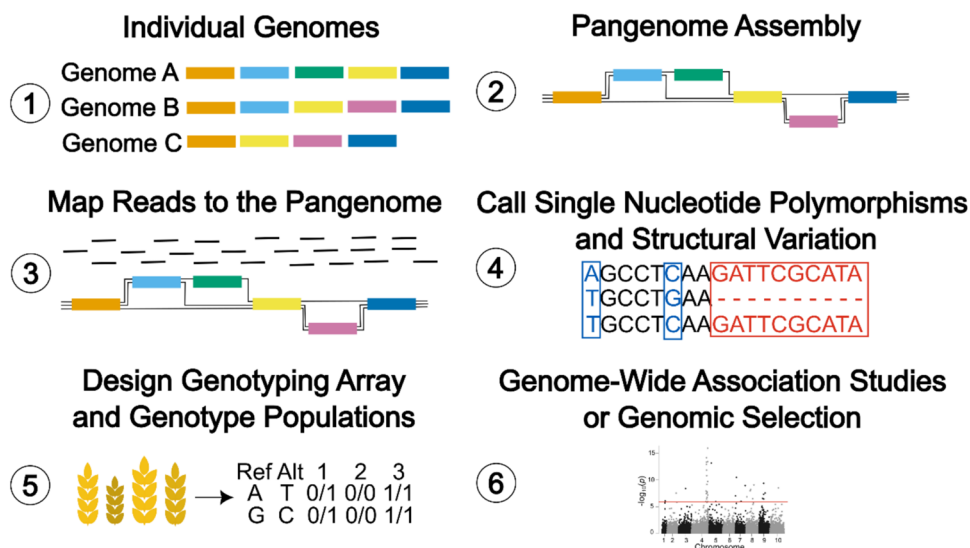
Table 2 (continued)

Genus	Species	Crop	Number of accessions	Pangenome type	Wild accessions	References
<i>Solanum</i>	<i>S. berthaultii</i> , <i>S. cardio-phyllum</i> , <i>S. ehrenbergii</i> , <i>S. infundibuliforme</i> , <i>S. kurtzianum</i> , <i>S. microdon-tum</i> , <i>S. okadae</i> , <i>S. polyad-enium</i> , <i>S. raphanifolium</i> , and <i>S. verrucosum</i> , and <i>S. violaceimarmoratum</i>	Potato	50	Graph	Yes	Zhu et al. 2025
<i>Solanum</i>	<i>S. lycopersicum</i> , <i>S. lycoper-sicoides</i> , <i>S. habrochaïtes</i> , <i>S. pennellii</i> , <i>S. chilense</i> , <i>S. peruvianum</i> , <i>S. cor-neliomulleri</i> , <i>S. neorickii</i> , <i>S. chmielewskii</i> , <i>S. pimpi-nellifolium</i> , and <i>S. galapa-gense</i>	Tomato	13	Graph	Yes	N. Li et al. 2023a
<i>Sorghum</i>	<i>S. propinquum</i> and <i>S. bicolor</i>	Sorghum	16	Graph	Yes	Tao et al. 2021
<i>Spinacia</i>	<i>S. oleracea</i> , <i>S. turkestanica</i> , and <i>S. tetrandra</i>	Spinach	13	Gene-based	Yes	She et al. 2024
<i>Vitis</i>	29 species	Grapevine	72	Graph	Yes	L. Guo et al. 2025a
<i>Vitis</i>	<i>V. acerifolia</i> , <i>V. aestivalis</i> , <i>V. arizonica</i> , <i>V. berlandieri</i> , <i>V. girdiana</i> , <i>V. monticola</i> , <i>V. mustangensis</i> , <i>V. riparia</i> , and <i>V. rupestris</i>	Grapevine	18	Graph	Yes	Cochetel et al. 2023
<i>Zea</i>	<i>Z. mays</i> , <i>Z. nicaraguensis</i> , <i>Z. luxurians</i> , <i>Z. diploperennis</i> , and <i>Z. perennis</i>	Maize	721	de novo and Graph	No	Gui et al. 2022

have been identified, they can be used to develop pangenome-based genotyping arrays which efficiently, accurately and inexpensively genotype a population and provide markers for PAB methods such as pangenome-wide association studies (pan-GWAS) and genomic selection (GS) (Fig. 1).

Current pangenome-based genotyping arrays include: Bna-Pan50T (Li et al. 2023b), a *Brassica napus* genotyping array with 54,765 markers for 588 accessions; rice pangenome genotyping array (RPGA) (Daware et al. 2023), including 271 rice accessions and 80 K markers; and *Triticum*

Fig. 1 A workflow for the application of pangenomes for crop improvement. Genome assemblies of the crop of interest are collected and assembled into a pangenome. Pangenome-based genomic markers are identified by mapping reads to the pangenome and calling single nucleotide polymorphisms and structural variation. Once pangenome markers have been identified, they can be used to develop pangenome-based genotyping arrays. Markers identified by pangenome-based genotyping arrays can then be used in genome-wide association studies and genomic selection



aestivum next generation array (TaNG) (BurrIDGE et al. 2024), a wheat genotyping array including 315 accessions and 43,372 markers.

Pangenome-based GWAS has been studied in several crops (Table 3). A study in sorghum (Tao et al. 2021) constructed a pangenome from 16 wild and domesticated accessions. The SNPs called from this were used in a GWAS analysis to identify 25 SNPs significantly associated with grain color. One of the most significant SNPs overlapped with a region of PAV that was associated with the known grain color gene *Yellow seed1*, identifying novel genetic variation associated with sorghum seed color that could not have been identified using single reference based GWAS alone. Chen and colleagues (2023) also found that pangenome-based GWAS could identify novel genetic variation for known causal genes. They used SNPs and SVs in a GWAS study of 43 agronomically important traits in broomcorn millet. The significantly associated SNPs and SVs overlapped. However, the significant SVs highlighted causal mutations associated with the traits of interest. For example, both significant SNPs and a significant SV were identified in *longmi008332*, a gene associated with inflorescence and seed color. The significant SV identified was an insertion that was found to be the causal mutation determining green or purple inflorescence. SV-based GWAS can be used together with SNP-based GWAS to examine genetic regions associated with agronomically important traits in greater detail to identify causal mutants and novel genetic variation. Pan-GWAS in tea identified both SNPs and SVs associated with leaf color, timing of bud flush, and flavor (Chen et al. 2023b). This study identified significant SVs overlapping with SNPs that were used to identify potential causal mutants, but they also identified significant SVs in novel genetic regions and found that only 9.1% of significant SVs overlapped with significant SNPs. These studies demonstrate the ability of pan-GWAS to not only identify potential causal mutations but to also identify novel sources of genetic variation that could be used to improve crops.

Recently, Yildiz et al. (2025) reported the first application of graphical pangenomics approaches for expression quantitative trait locus (eQTL) analysis in crops. A pangenome graph was built and used as reference for both genotyping and transcript expression quantification of 100 winter *Brassica napus* lines. Pangenome graph-based transcript quantification proved more accurate than linear-based, with transcripts with medium/low correlation between approaches being significantly overrepresented in SNPs and SVs. Noticeably, approximately a third of the transcripts associated with eQTL-SVs had no significant association with SNPs, highlighting the importance of including SVs in association studies and the added value of pangenome graph-based eQTL analyses. Altogether, this work resulted in the identification of hundreds of gene proximal eQTL-SVs, with

several variants affecting stress response- and morphogenesis- related transcripts levels, pointing to a potential role of SV-driven expression variation in phenotypic diversity and trait fine-tuning in rapeseed.

The use of pangenome-based markers in GS has been found to increase trait prediction accuracy in several crops (Table 3). A GS study in pearl millet used both SNPs and SVs to identify breeding candidates for 120 agronomically important phenotypes (Yan et al. 2024). The SNP-based GS identified 9 breeding candidates while the SV-based GS identified 4. The SV-based GS captured genetic diversity not identified by SNPs and had higher prediction accuracy, demonstrating the benefit of using both SNPs and SVs in GS. Studies in tomatoes and grapevine found a 24% and 22.78% increase in estimated heritability when using a pangenome compared to a single reference genome, respectively (Liu et al. 2024b; Zhou et al. 2022). This increase in heritability due to the inclusion of SVs was credited for improving the power of GWAS to identify significantly associated markers and for the increase of trait prediction accuracy. A similar study in spinach supports the conclusion that capturing missing heritability with SVs improves trait prediction accuracy (She et al. 2024) and that PAB can improve the accuracy and efficiency of breeding programs compared to single genome reference based assisted breeding.

Genome editing is an alternative to conventional breeding methods for introducing SVs into crops. Sun et al. (2024) used CRISPR/Cas9 to successfully insert, delete, and translocate SVs of up to 30 Mb in rice. Similarly, Y. Lu et al. (2021) improved herbicide resistance in rice without affecting other agronomic traits, by generating a 911 kb inversion on chromosome 1 and a 338 kb duplication on chromosome 2. These studies highlight how genome editing can use SVs for crop improvement.

Pan-omics

Pan-epigenomics is an approach seeking to characterize epigenetic variation within a species, complementing pangenomics and pan-transcriptomics with the study of epigenetic modifications such as DNA methylation, histone modifications, and chromatin accessibility. These modifications play an important role in regulating gene expression, transposable element activity, and developmental processes, and have been shown to contribute to intraspecific diversity and several crop traits, including stress tolerance, flowering behavior and yield (Crisp et al. 2016; Eichten et al. 2013; Guo et al. 2023; Lieberman-Lazarovich et al. 2022; Vafadarshamasbi et al. 2022; Xu et al. 2020, 2019)(See Crisp et al. 2016; Lieberman-Lazarovich et al. 2022; Vafadarshamasbi et al. 2022 for reviews on the topic). Complementarily, structural

Table 3 Pangenome assisted breeding (PAB) studies in crops. The PAB studies include genome-wide association studies (GWAS) and genomic selection (GS) studies. These PAB studies used single nucleotide polymorphisms (SNPs) and/or structural variation (SV) markers to identify genomic regions and individuals associated with an agronomically important trait

Species	Crop	Pangenome type	Number of accessions	PAB method	Marker type	Trait studied	References
<i>Brassica napus</i>	Rapeseed	de novo	8	GWAS	SNPs and SVs	Silique length, seed weight, and flowering time	Song et al. 2020
<i>Brassica napus</i>	Rapeseed	Graph	57–100	eQTL	SNPs and SVs	Expression	Yildiz et al. 2025
<i>Camellia sinensis</i>	Tea	Graph	22	GWAS	SNPs and SVs	Leaf colour, timing of bud flush, and flavour-related chemistries	S. Chen et al. 2023b
<i>Cicer arietinum</i>	Chickpea	Iterative mapping and assembly	3202	GWAS	SNPs	16 traits	Varshney et al. 2021
<i>Cicer arietinum</i>	Chickpea	Iterative mapping and assembly	3202	GS	SNPs	9 traits	Varshney et al. 2021
<i>Citrullus</i>	Watermelon	Graph	28	GWAS	SNPs and SVs	Sugar content and flesh colour	Yilin Zhang et al. 2024b
<i>Cucumis sativus</i>	Cucumber	Graph	12	GWAS	SNPs and SVs	Female flower rate on a primary branch, fruit spine/wart density, and branch number	H. Li et al. 2022
<i>Glycine max</i>	Soybean	Iterative mapping and assembly	915	GWAS	SNPs	Flowering time	Mohamedikbal et al. 2024
<i>Lactuca</i>	Lettuce	Iterative mapping and assembly	474	GWAS	SVs	Disease resistance	van Workum et al. 2024
<i>Oryza sativa</i>	Rice	Graph	33	GWAS	SNPs and SVs	Leaf senescence	Qin et al. 2021
<i>Oryza sativa</i>	Rice	Iterative mapping and assembly	3010	GWAS	SNPs	Grain size/weight traits	Daware et al. 2023
<i>Panicum mili-aceum</i>	Broomcorn millet	Graph	32	GWAS	SNPs and SVs	43 traits	J. Chen et al. 2023a
<i>Pennisetum glaucum</i>	Pearl millet	Graph	11	GS	SNPs and SVs	120 traits	Yan et al. 2024
<i>Setaria italica</i>	Foxtail millet	Graph	110	GWAS	SNPs and SVs	68 traits	Q. He et al. 2023
<i>Setaria italica</i>	Foxtail millet	Graph	110	GS	SNPs and SVs	68 traits	Q. He et al. 2023
<i>Solanum</i>	Potato	Graph	50	GWAS	SVs	Metabolite content	Zhu et al. 2025
<i>Solanum lycopersicum</i>	Tomato	Graph	31	GWAS	SNPs and SVs	Fruit soluble solids content	Zhou et al. 2022
<i>Solanum</i>	Tomato	Graph	13	GWAS	SNPs	Flavour-related metabolite traits	N. Li et al. 2023a
<i>Sorghum</i>	Sorghum	Graph	16	GWAS	SNPs	Grain colour	Tao et al. 2021
<i>Sorghum bicolor</i>	Sorghum	Graph	24	GS	SNPs	Height, brix, juice weight, leaf weight, ear-liness, stem weight, and grain yield	Jensen et al. 2020
<i>Spinacia</i>	Spinach	Gene-based	13	GWAS	SNPs and SVs	sex	She et al. 2024
<i>Spinacia</i>	Spinach	Gene-based	13	GS	SNPs and SVs	17 traits	She et al. 2024
<i>Triticum aestivum</i>	Wheat	k-mer-based	139	GWAS	SVs	Seed-storage proteins	Z. Zhang et al. 2024d
<i>Triticum aestivum</i>	Wheat	k-mer-based	139	GS	SVs	Seed-storage proteins	Z. Zhang et al. 2024d

Table 3 (continued)

Species	Crop	Pangenome type	Number of accessions	PAB method	Marker type	Trait studied	References
<i>Vitis vinifera</i>	Grapevine	Graph	29	GWAS	SNPs and SVs	29 traits	Z. Liu et al. 2024b
<i>Vitis</i>	Grapevine	Graph	18	GWAS	SNPs and SVs	Chloride exclusion	Cochetel et al. 2023
<i>Zea</i>	Maize	Graph	507	GWAS	SNPs and SVs	Agronomic traits, metabolites, and protein contents	Gui et al. 2022

variations can disrupt gene structures, alter both near and distal regulatory sequences, or reposition transposable elements, thereby affecting local and global epigenetic landscapes (see Noshay and Springer 2021; Zhang et al. 2018 for literature reviews). This non-coding variation can be the basis for elucidating molecular mechanisms and regulatory networks in plants and also provide novel sources for fine-tuning crop genetic improvement (see Crisp et al. 2022; Ganguly et al. 2022; King et al. 2010; Lieberman-Lazarovich et al. 2022; Zhi and Chang 2021 for reviews discussing the topic).

While plant pan-epigenomic studies are only at their infancy due to the previously prohibitive costs of assessing genome-wide epigenetic features for multiple individuals, recent advancements such as single-molecule real-time sequencing and nanopore-based methylation detection have enabled high-resolution epigenomic profiling. Notably, Noshay and colleagues (Noshay et al. 2021) investigated the variability of both chromatin accessibility and DNA methylation across four maize inbred lines with high-quality reference genomes available. This study revealed that, while only 5% of unmethylated regions (UMRs) were polymorphic between genotypes, more than 70% of the conserved UMRs exhibited incomplete overlaps, even in the absence of underlying sequence variation. Similarly, the majority of accessible chromatin regions (ACRs) were conserved between genotypes, but almost 29% of accessible UMRs in one accession were found unmethylated and/or inaccessible in the other genotypes. This previously underexplored epigenetic variability could be an important contributor to intraspecific differences with functional consequences.

Recently, Zhao et al. (2024) generated population-wide DNA methylation data and integrated this with both genetic polymorphism and transcriptomic data for 207 cotton accessions. Association analyses identified 5.4 million cis-methylation quantitative trait loci (cis-meQTLs), and 5000 cis-expression quantitative trait methylations (cis-eQTM), with approximately 36% of cis-eQTM genes not found associated with genic variations. Notably, one cis-eQTM gene that was not identified by GWAS alone was validated via gene editing, proving its role in fiber development. These

results establish the potential of DNA methylation variation for identifying novel opportunities for crop improvement.

Pan-transcriptomics is another emerging field, with the development of sequencing technology making multi-tissue and multi-environmental sampling more accessible. A pioneering work from 2014 investigated a maize pan-transcriptome in a panel of 503 inbred lines at the seedling stage, enabling the identification of transcript abundance and transcript PAV associated with the transition between juvenile to adult growth and with flowering time (Hirsch et al. 2014). More recently, a pan-transcriptome of 20 barley individuals was constructed using short- and long-read RNA-sequencing along with Iso-seq data (Guo et al. 2025b). The barley pan-transcriptome improved mapping rate and found more transcripts per gene compared to a single-genome reference. Analysis of gibberellic acid (GA) metabolism, using the pan-transcriptome, found differential expression of GA genes including the overexpression of *GA2ox3*, which was predicted to improve yield under environmental stress (Guo et al. 2025b). A rice pan-transcriptome found an increased number of multiple splicing isoforms compared to a single-genome reference (Zhong et al. 2024). Zhong and colleagues (2024) constructed the pan-transcriptome from a variety of cultivars and cold stress conditions and identified cultivar- and cold stress-specific transcripts and alternative splicing sites. It was found that differentially expressed genes and alternative splicing-regulated gene expression were associated with cold stress tolerance, which varied between cultivars (Zhong et al. 2024). These results could inform breeders of which cultivars are better adapted to cold stress.

Alternative splicing (AS) events are known but largely unexplored sources of intraspecific diversity, with an estimated 20% of genes having one or more alternative transcript isoforms in plants (see Barbazuk et al. 2008 for a review). To our knowledge, there are no pan-multi-omics studies that focus on alternative splicing events in plants, despite studies finding DNA methylation and the epigenome play an important role in affecting alternative splicing events for agronomically important traits under environmental stress conditions (Qian et al. 2019; Wang et al. 2021). Nevertheless, novel tools are being developed

to investigate AS events using pangenomes, namely the *pantas* toolkit that combines a genomic variation graph with short-read RNA sequencing data to annotate the graph pangenome with alternative splice event information (Ciccolella et al. 2024). *Pantas* outperformed alternative splicing site identification tools based on a single-genome references (Ciccolella et al. 2024). This tool can combine genomic and transcriptomic data to identify alternative splice sites that are associated with agronomically important traits. Future studies should develop pangenome methods that can incorporate transcriptomic and epigenomic data alongside genomic data, such as *pantas*. Additionally, further research should be done on the effect of the transcriptome and epigenome on alternative splicing events associated with agronomically important traits.

Given these promising case studies pan-transcriptome and pan-epigenome atlases, profiling gene expression and epigenetic marks across a wide range of genetic backgrounds with high-quality reference genomes, could prove instrumental (see Huang et al. 2022; Jayakodi et al. 2021 for reviews on the topic). These atlases will be instrumental in deciphering the interplay between genetic variation, epigenetic modifications, and environmental factors in shaping phenotypic diversity.

Developing user-friendly pangenome browsers with integrated transcriptomics and epigenetic data will be useful for making this complex information accessible to researchers and breeders. Existing platforms such as Panache (Durant et al. 2021) and PPanG (Liu et al. 2024a) visualize pangenomes alongside individual sequences and tracks with gene annotation, coordinates, and alignments. Visualization tools such as these could be adapted to display additional tracks for UMRs and ACRs. Users could then select different environmental conditions and visualize methylation changes for each individual. Additionally, for epigenomic markers found to be associated with important traits, functional annotation could be displayed. To our knowledge, no pangenome browsers have integrated visualization for epigenetic information; however, these visualization tools would support comparative epigenomics, allowing users to track epigenetic changes across different accessions and conditions. Such resources will enhance the application of pan-epigenomics in precision breeding, supporting the selection of genotypes with desirable epigenetic traits for improved yield, stress tolerance, and adaptability.

Overall, pan-epigenomics and pan-transcriptomics represent promising frontiers in plant genomics, expanding our understanding on how genetic epigenetic, and transcriptomic variation collectively shape phenotypes. By expanding our understanding of epigenetic and transcriptomic diversity within and between species, these fields have the potential to revolutionize crop improvement, ecological adaptation studies, and functional genomics research.

With the growth of multi-omic studies, pangenomes could be combined with other omics such as metabolomics and proteomics to avoid reference bias also in these investigations. A PAV-based metabolomics-GWAS in potatoes identified 9321 SVs associated with 1258 distinct metabolites (Zhu et al. 2025). These metabolites affected agronomically important traits such as starch and flavonoid content and were not identified by SNP-based GWAS. A combined genomic and pan-metabolic study in cassava found that genetic and leaf metabolite subgroups correlated with eco-geological regions, while root metabolite clades did not (Perez-Fons et al. 2023). In addition, genetic subgroups correlated with agronomically important traits such as dry matter content, cooking time, and cyanide content but not to whitefly tolerance, while leaf metabolites did correlate with whitefly tolerance. These studies highlight how genomics and metabolomics can identify different sources of variation. Combining pangenomics with other omics generates a more complete picture of crop systems and can be used to identify the causal genetic/epigenetic variants associated with agronomically important traits. Currently, there are pangenome-based multi-omics databases for key crops such as *Brassica napus* (Cui et al. 2023), cereals such as oat and barley (Yao et al. 2022), and potatoes (Zhu et al. 2025). *Asteraceae* multi-omics information resource (AMIR) is a database containing 74 *Asteraceae* species and corresponding genome, pangenome, functional annotation, variants, ortholog, and tissue specific transcriptome data (Liu et al. 2025). AMIR additionally includes bioinformatic tools such as BLAST and Gene Ontology to streamline bioinformatic analyses within *Asteraceae* crops. Liu and colleagues (2025) demonstrated the use of AMIR by searching their pangenome for a known agronomically important gene, confirming the gene was a core gene, identifying homologous genes, and constructing a phylogenetic tree of homologs. More databases such as these could be collated and made public to foster global collaboration and innovation.

Conclusion

The decreasing cost and increasing quality of sequencing data is driving the development of more pangenomes and expanding the scope of comparative pangenomics. Previously, genes were considered lost if they were missing from a single reference, however now we can measure the frequency of gene presence in a population. This frequency measure can be compared within and between species to help us understand selection at both the individual and population scale. Several minor crops and many wild plant species have not benefitted from pangenome analysis. Minor crops and wild relatives are an excellent source of genetic variation that can be used to improve the major crop species

and also support the domestication of new species to expand the diversity of food crops. Crop pangenomics is a rapidly developing research area that can improve our understanding of the genetic diversity underlying agronomic traits; this knowledge can be used to improve crops through pangenomics assisted breeding and genome editing.

Acknowledgements This research was carried out while TRM was in receipt of an Australian Government Research Training Program Stipend at The University of Western Australia. We acknowledge support from the Australian Research Council project LP230100351.

Author Contributions All authors performed the literature search, drafted, and revised the review. All authors read and approved the final manuscript.

Funding Open Access funding enabled and organized by CAUL and its Member Institutions. DE acknowledges funding from the Australia Research Council (Project LP230100351). VK, RJS and AAG were supported by DFG grant 469336000 from the German Research Society (DFG) for the International Research Training Group IRTG 2843 “Accelerating Crop Genetic Gain”. This research was supported by the LOEWE Start Professorship from the Hessian Ministry of Higher Education, Research, Science and the Arts to and Sofja Kovalevskaja Award from the Humboldt Foundation to AAG.

Declarations

Conflict of interest The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Alonge M, Wang X, Benoit M, Soyk S, Pereira L, Zhang L, Suresh H, Ramakrishnan S, Maumus F, Ciren D, Levy Y, Harel TH, Shalev-Schlosser G, Amsellem Z, Razifard H, Caicedo AL, Tieman DM, Klee H, Kirsche M, Aganezov S, Ranallo-Benavidez TR, Lemmon ZH, Kim J, Robitaille G, Kramer M, Goodwin S, McCombie WR, Hutton S, Van Eck J, Gillis J, Eshed Y, Sedlazeck FJ, van der Knaap E, Schatz MC, Lippman ZB (2020) Major impacts of widespread structural variation on gene expression and crop improvement in tomato. *Cell* 182:145–161.e123. <https://doi.org/10.1016/j.cell.2020.05.021>
- Andolfo G, D'Agostino N, Frusciante L, Ercolano MR (2021) The tomato interspecific NB-LRR gene arsenal and its impact on breeding strategies. *Genes* 12:174. <https://doi.org/10.3390/genes12020184>
- Andreace F, Lechat P, Dufresne Y, Chikhi R (2023) Comparing methods for constructing and representing human pangenome graphs. *Genome Biol* 24:Article 274. <https://doi.org/10.1186/s13059-023-03098-2>
- Barbazuk WB, Fu Y, McGinnis KM (2008) Genome-wide analyses of alternative splicing in plants: opportunities and challenges. *Genome Res* 18:1381–1392. <https://doi.org/10.1101/gr.053678.106>
- Batley J, Edwards D (2016) The application of genomics and bioinformatics to accelerate crop improvement in a changing climate. *Curr Opin Plant Biol* 30:78–81. <https://doi.org/10.1016/j.pbi.2016.02.002>
- Battlay P, Hendrickson BT, Mendez-Reneau JI et al. (2024) Structural variants underlie parallel adaptation following global invasion. *bioRxiv*. <https://www.biorxiv.org/content/>, <https://doi.org/10.1101/2024.07.09.602765v1.full.pdf>
- Bayer PE, Hurgobin B, Golicz AA, Chan CK, Yuan Y, Lee H, Renton M, Meng J, Li R, Long Y, Zou J, Bancroft I, Chalhoub B, King GJ, Batley J, Edwards D (2017) Assembly and comparison of two closely related *Brassica napus* genomes. *Plant Biotechnol J* 15:1602–1610. <https://doi.org/10.1111/pbi.12742>
- Bayer PE, Peterleit J, Durant É, Monat C, Rouard M, Hu H, Chapman B, Li C, Cheng S, Batley J, Edwards D (2022a) Wheat Panache: a pangenome graph database representing presence–absence variation across sixteen bread wheat genomes. *Plant Genome* 15:e20221. <https://doi.org/10.1002/tpg2.20221>
- Bayer PE, Valliyodan B, Hu H, Marsh JI, Yuan Y, Vuong TD, Patil G, Song Q, Batley J, Varshney RK, Lam H, Edwards D, Nguyen HT (2022b) Sequencing the USDA core soybean collection reveals gene loss during domestication and breeding. *Plant Genome* 15:e20109. <https://doi.org/10.1002/tpg2.20109>
- Blanc G, Wolfe KH (2004) Functional divergence of duplicated genes formed by polyploidy during *Arabidopsis* evolution. *Plant Cell* 16:1679–1691. <https://doi.org/10.1105/tpc.021410>
- Bozan I, Achakkagari SR, Anglin NL, Ellis D, Tai HH, Stromvik MV (2023) Pangenome analyses reveal impact of transposable elements and ploidy on the evolution of potato species. *Natl Acad Sci* 120:e2211117120–e2211117120. <https://doi.org/10.1073/pnas.2211117120>
- Bradbury PJ, Casstevens T, Jensen SE, Johnson LC, Miller ZR, Monier B, Romay MC, Song B, Buckler ES (2022) The practical haplotype graph, a platform for storing and using pangenomes for imputation. *Bioinformatics* 38:3698–3702. <https://doi.org/10.1093/bioinformatics/btac410>
- Burridge AJ, Winfield M, Przewieslik-Allen A, Edwards KJ, Siddique I, Barral-Arca R, Griffiths S, Cheng S, Huang Z, Feng C, Dreisigacker S, Bentley AR, Brown-Guedira G, Barker GL (2024) Development of a next generation SNP genotyping array for wheat. *Plant Biotechnol J* 22:2235–2247. <https://doi.org/10.1111/pbi.14341>
- Chalhoub B, Denoeud F, Liu S, Parkin IAP, Tang H, Wang X, Chiquet J, Belcram H, Tong C, Samans B, Corréa M, Da Silva C, Just J, Falentin C, Koh CS, Le Clainche I, Bernard M, Bento P, Noel B, Labadie K, Alberti A, Charles M, Arnaud D, Guo H, Daviaud C, Alamery S, Jabbari K, Zhao M, Edger PP, Chelaifa H, Tack D, Lassalle G, Mestiri I, Schnel N, Le Paslier M-C, Fan G, Renault V, Bayer PE, Golicz AA, Manoli S, Lee T-H, Thi VHD, Chalabi S, Hu Q, Fan C, Tollenaere R, Lu Y, Battail C, Shen J, Sidebottom CHD, Wang X, Canaguier A, Chauveau A, Bérard A, Deniot G, Guan M, Liu Z, Sun F, Lim YP, Lyons E, Town CD, Bancroft I, Wang X, Meng J, Ma J, Pires JC, King GJ, Brunel D, Delourme R, Renard M, Aury J-M, Adams KL, Batley J, Snowdon RJ, Tost J, Edwards D, Zhou Y, Hua W, Sharpe AG, Paterson AH, Guan C, Wincker P (2014) Early allopolyploid evolution in the post-neolithic *Brassica napus* oilseed genome. *Science* 346:950–953. <https://doi.org/10.1126/science.1253435>
- Chawla HS, Lee HT, Gabur I, Tamilselvan-Nattar-Amutha S, Obermeier C, Schiessl SV, Song JM, Liu K, Guo L, Parkin IAP,

- Snowdon RJ (2020) Long-read sequencing reveals widespread intragenic structural variants in a recent allopolyploid crop plant. *Plant Biotech J*. <https://doi.org/10.1111/pbi.13456>
- Chen J, Liu Y, Liu M, Guo W, Wang Y, He Q, Chen W, Liao Y, Zhang W, Gao Y, Dong K, Ren R, Yang T, Zhang L, Qi M, Li Z, Zhao M, Wang H, Wang J, Qiao Z, Li H, Jiang Y, Liu G, Song X, Deng Y, Li H, Yan F, Dong Y, Li Q, Li T, Yang W, Cui J, Wang H, Zhou Y, Zhang X, Jia G, Lu P, Zhi H, Tang S, Diao X (2023a) Pangenome analysis reveals genomic variations associated with domestication traits in broomcorn millet. *Nat Genet* 55:2243–2254. <https://doi.org/10.1038/s41588-023-01571-z>
- Chen S, Wang P, Kong W, Chai K, Zhang S, Yu J, Wang Y, Jiang M, Lei W, Chen X, Wang W, Gao Y, Qu S, Wang F, Wang Y, Zhang Q, Gu M, Fang K, Ma C, Sun W, Ye N, Wu H, Zhang X (2023b) Gene mining and genomics-assisted breeding empowered by the pangenome of tea plant *Camellia sinensis*. *Nat Plants* 9:1986–1999. <https://doi.org/10.1038/s41477-023-01565-z>
- Cheng H, Concepcion GT, Feng X, Zhang H, Li H (2021) Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nat Methods* 18:170–175. <https://doi.org/10.1038/s41592-020-01056-5>
- Ciccolella S, Cozzi D, Della Vedova G, Kuria SN, Bonizzoni P, Denti L (2024) Differential quantification of alternative splicing events on spliced pangenome graphs. *PLoS Comput Biol* 20:e1012665. <https://doi.org/10.1371/journal.pcbi.1012665>
- Cochetel N, Minio A, Guarracino A, Garcia JF, Figueroa-Balderas R, Massonnet M, Kasuga T, Londo JP, Garrison E, Gaut BS, Cantu D (2023) A super-pangenome of the North American wild grape species. *Genome Biol* 24:290–290. <https://doi.org/10.1186/s13059-023-03133-2>
- Crisp PA, Ganguly D, Eichten SR, Borevitz JO, Pogson BJ (2016) Reconsidering plant memory: intersections between stress recovery, RNA turnover, and epigenetics. *Sci Adv* 2:e1501340. <https://doi.org/10.1126/sciadv.1501340>
- Crisp PA, Bhatnagar-Mathur P, Hundleby P, Godwin ID, Waterhouse PM, Hickey LT (2022) Beyond the gene: epigenetic and cis-regulatory targets offer new breeding potential for the future. *Curr Opin Biotechnol* 73:88–94. <https://doi.org/10.1016/j.copbio.2021.07.008>
- Cui X, Hu M, Yao S, Zhang Y, Tang M, Liu L, Cheng X, Tong C, Liu S (2023) BnaOmics: a comprehensive platform combining pan-genome and multi-omics data from *Brassica napus*. *Plant Commun* 4:100609–100609. <https://doi.org/10.1016/j.xplc.2023.100609>
- Daware A, Malik A, Srivastava R, Das D, Ellur RK, Singh AK, Tyagi AK, Parida SK (2023) Rice pangenome genotyping array: an efficient genotyping solution for pangenome-based accelerated genetic improvement in rice. *Plant J* 113:26–46. <https://doi.org/10.1111/tpj.16028>
- Ding B, Hu H, Cao Y, Xu R, Lin Y, Tulq M, Song Y, He G, Han Y, Guo H, Qiao J, Zhao J, Feng X, Yang S, Guo X, Varshney RK, Li L (2024) Pear genomes display significant genetic diversity and provide novel insights into the fruit quality traits differentiation. *Hortic Plant J* 10:1274–1290. <https://doi.org/10.1016/j.hpj.2024.05.005>
- Durant É, Sabot F, Conte M, Rouard M (2021) Panache: A web browser-based viewer for linearized pangenomes. *Bioinformatics* 37:4556–4558. <https://doi.org/10.1093/bioinformatics/btab688>
- Eichten SR, Briskine R, Song J, Li Q, Swanson-Wagner R, Hermanson PJ, Waters AJ, Starr E, West PT, Tiffin P, Myers CL, Vaughn MW, Springer NM (2013) Epigenetic and genetic influences on DNA methylation variation in maize populations. *Plant Cell*. <https://doi.org/10.1105/tpc.113.114793>
- Feng JW, Pidon H, Cuacos M, Lux T, Himmelbach A, Haghi R, Fuchs J, Haberer G, Kuo YT, Guo Y, Jayakodi M, Toegelová H, Harpke D, Knauf M, Fiebig A, Maruschewski M, Ronen M, Sharon A, Šimková H, Mayer KFX, Spannagl M, Kumlehn J, Heckmann S, Houben A, Blattner FR, Stein N, Mascher M (2025) A haplotype-resolved pangenome of the barley wild relative *Hordeum bulbosum*. *Nature*. <https://doi.org/10.1038/s41586-025-09270-x>
- Fuentes RR, Chebotarov D, Duitama J, Smith S, De la Hoz JF, Mohiyuddin M et al (2019) Structural variants in 3000 rice genomes. *Genome Res* 29:870–880. <https://doi.org/10.1101/gr.241240.118>
- Gabur I, Chawla HS, Lopisso DT, von Tiedemann A, Snowdon RJ, Obermeier C (2020) Gene presence-absence variation associates with quantitative *Verticillium longisporum* disease resistance in *Brassica napus*. *Sci Rep* 10:4131. <https://doi.org/10.1038/s41598-020-61228-3>
- Ganguly DR, Hickey LT, Crisp PA (2022) Harnessing genetic variation at regulatory regions to fine-tune traits for climate-resilient crops. *Mol Plant*. <https://doi.org/10.1016/j.molp.2021.12.011>
- Garrison E, Sirén J, Novak AM, Hickey G, Eizenga JM, Dawson ET, Jones W, Garg S, Markello C, Lin MF, Paten B, Durbin R (2018) Variation graph toolkit improves read mapping by representing genetic variation in the reference. *Nat Biotechnol* 36:875–879. <https://doi.org/10.1038/nbt.4227>
- Garrison E, Guarracino A, Heumos S, Villani F, Bao Z, Tattini L, Hagmann J, Vorbrugg S, Marco-Sola S, Kubica C, Ashbrook DG, Thorell K, Rusholme-Pilcher RL, Liti G, Rudbeck E, Golicz AA, Nahnsen S, Yang Z, Mwaniki MN, Nobrega FL, Wu Y, Chen H, de Ligt J, Sudmant PH, Huang S, Weigel D, Soranzo N, Colonna V, Williams RW, Prins P (2024) Building pangenome graphs. *Nat Methods* 21:2008–2012. <https://doi.org/10.1038/s41592-024-02430-3>
- Goel M, Sun H, Jiao W-B, Schneeberger K (2019) SyRI: finding genomic rearrangements and local sequence differences from whole-genome assemblies. *Genome Biol* 20:277. <https://doi.org/10.1186/s13059-019-1911-0>
- Golicz AA, Martinez PA, Zander M, Patel DA, Van De Wouw AP, Visendi P, Fitzgerald TL, Edwards D, Batley J (2015) Gene loss in the fungal canola pathogen *Leptosphaeria maculans*. *Funct Integr Genomics* 15:189–196. <https://doi.org/10.1007/s10142-014-0412-1>
- Golicz AA, Bayer PE, Barker GC, Edger PP, Kim H, Martinez PA, Chan CKK, Severn-Ellis A, McCombie WR, Parkin IAP, Paterson AH, Pires JC, Sharpe AG, Tang H, Teakle GR, Town CD, Batley J, Edwards D (2016) The pangenome of an agronomically important crop plant *Brassica oleracea*. *Nat Commun* 7:13390–13398. <https://doi.org/10.1038/ncomms13390>
- Guarracino A, Heumos S, Nahnsen S, Prins P, Garrison E (2022) ODGI: understanding pangenome graphs. *Bioinformatics* 38:3319–3326. <https://doi.org/10.1093/bioinformatics/btac308>
- Gui S, Wei W, Jiang C, Luo J, Chen L, Wu S, Li W, Wang Y, Li S, Yang N, Li Q, Fernie AR, Yan J (2022) A pan-*Zea* genome map for enhancing maize improvement. *Genome Biol* 23:Article 178. <https://doi.org/10.1186/s13059-022-02742-7>
- Guo H, Cao P, Wang C, Lai J, Deng Y, Li C, Hao Y, Wu Z, Chen R, Qiang Q, Fernie AR, Yang J, Wang S (2023) Population analysis reveals the roles of DNA methylation in tomato domestication and metabolic diversity. *Sci China Life Sci*. <https://doi.org/10.1007/s11427-022-2299-5>
- Guo L, Wang X, Ayhan DH, Rhaman MS, Yan M, Jiang J, Wang D, Zheng W, Mei J, Ji W, Jiao J, Chen S, Sun J, Yi S, Meng D, Wang J, Bhuiyan MN, Qin G, Guo L, Yang Q, Zhang X, Sun H, Liu C, Deng XW, Ye W (2025a) Super pangenome of *Vitis** empowers identification of downy mildew resistance genes for grapevine improvement. *Nat Genet* 57:741–753. <https://doi.org/10.1038/s41588-025-02111-7>

- Guo W, Schreiber M, Marosi VB, Bagnaresi P, Jørgensen ME, Braune KB, Chalmers K, Chapman B, Dang V, Dockter C, Fiebig A, Fincher GB, Fricano A, Fuller J, Haaning A, Haberer G, Himmelbach A, Jayakodi M, Jia Y, Kamal N, Langridge P, Li C, Lu Q, Lux T, Mascher M, Mayer KFX, McCallum N, Milne L, Muehlbauer GJ, Nielsen MTS, Padmarasu S, Pidas PR, Pillen K, Pozniak C, Rasmussen MW, Sato K, Schmutzer T, Scholz U, Schüler D, Šimková H, Skadhauge B, Stein N, Thomsen NW, Voss C, Wang P, Wonneberger R, Zhang X-Q, Zhang G, Cattivelli L, Spannagl M, Bayer M, Simpson C, Zhang R, Waugh R (2025b) A barley pan-transcriptome reveals layers of genotype-dependent transcriptional complexity. *Nat Genet* 57:129. <https://doi.org/10.1038/s41588-024-02069-y>
- He Z, Ji R, Havlickova L, Wang L, Li Y, Lee HT, Song J, Koh C, Yang J, Zhang M, Parkin IAP, Wang X, Edwards D, King GJ, Zou J, Liu K, Snowdon RJ, Banga SS, Machackova I, Bancroft I (2021) Genome structural evolution in *Brassica* crops. *Nat Plants* 7:757–765. <https://doi.org/10.1038/s41477-021-00928-8>
- He Q, Tang S, Zhi H, Chen J, Zhang J, Liang H, Alam O, Li H, Zhang H, Xing L, Li X, Zhang W, Wang H, Shi J, Du H, Wu H, Wang L, Yang P, Xing L, Yan H, Song Z, Liu J, Wang H, Tian X, Qiao Z, Feng G, Guo R, Zhu W, Ren Y, Hao H, Li M, Zhang A, Guo E, Yan F, Li Q, Liu Y, Tian B, Zhao X, Jia R, Feng B, Zhang J, Wei J, Lai J, Jia G, Purugganan M, Diao X (2023) A graph-based genome and pan-genome variation of the model plant *Setaria**. *Nat Genet* 55:1232–1242. <https://doi.org/10.1038/s41588-023-01423-w>
- Heller D, Vingron M (2020) SVIM-asm: structural variant detection from haploid and diploid genome assemblies. *Bioinformatics* 6:5519–5521. <https://doi.org/10.1093/bioinformatics/btaa1034>
- Heuberger M, Bernasconi Z, Said M, Jung E, Herren G, Widrig V, Šimková H, Keller B, Sánchez-Martín J, Wicker T (2024) Analysis of a global wheat panel reveals a highly diverse introgression landscape and provides evidence for inter-homoeologue chromosomal recombination. *Theor Appl Genet* 137:236. <https://doi.org/10.1007/s00122-024-04721-x>
- Hickey G, Monlong J, Eblor J, Novak AM, Eizenga JM, Gao Y, Marschall T, Li H, Paten B (2024) Pangenome graph construction from genome alignments with Minigraph-Cactus. *Nat Biotechnol* 42:663–673. <https://doi.org/10.1038/s41587-023-01793-w>
- Hirsch CN, Foerster JM, Johnson JM, Sekhon RS, Muttoni G, Vaillancourt B, Peñagaricano F, Lindquist E, Pedraza MA, Barry K, de Leon N, Kaeppler SM, Buell CR (2014) Insights into the maize pan-genome and pan-transcriptome. *Plant Cell* 26:121–135. <https://doi.org/10.1105/tpc.113.119982>
- Holt C, Yandell M (2011) MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinform* 12:491–491. <https://doi.org/10.1186/1471-2105-12-491>
- Huang Y, Liu Y, Liu C, Birchler JA, Han F (2022) Prospects and challenges of epigenomics in crop improvement. *Gene Genom* 44:251–257. <https://doi.org/10.1007/s13258-021-01187-9>
- Hurgobin B, Golicz AA, Bayer PE, Chan CK, Tirnaz S, Dolatabadian A, Schiessl SV, Samans B, Montenegro JD, Parkin IAP, Pires JC, Chalhoub B, King GJ, Snowdon RJ, Batley J, Edwards D (2017) Homoeologous exchange is a major cause of gene presence/absence variation in the amphidiploid *Brassica napus*. *Plant Biotech J* 16:1265–1274. <https://doi.org/10.1111/pbi.12791>
- Jayakodi M, Schreiber M, Stein N, Mascher M (2021) Building pan-genome infrastructures for crop plants and their use in association genetics. *DNA Res*. <https://doi.org/10.1093/dnares/dsaa030>
- Jayakodi M, Golicz AA, Kreplak J, Fechet L, Angra D, Bednář P, Bornhofen E, Zhang H, Bousageon R, Kaur S, Cheung K, Čížková J, Gundlach H, Hallab A, Imbert B, Keeble-Gagnere G, Koblížková K, Koblová L, Krejčí P, Mouritzen T, Neumann P, Nadziejka M, Nielsen L, Novak P, Orabi J, Padmarasu S, Robertson-Shersby-Harvie S, Robledillo L, Schiemann A, Taniskanen J, Törönen P, Warsame A, Wittenberg A, Himmelbach A, Aubert G, Emmanuel Courty P, Doležel J, Holm L, Janss L, Khazaei H, Macas J, Mascher M, Smýkal P, Snowdon R, Stein N, Stoddard F, Stougaard J, Tayeh N, Torres A, Usadel B, Schubert I, O'Sullivan D, Schulman A, Andersen S (2023) The giant diploid faba genome unlocks variation in a global protein crop. *Nature* 615:652–659. <https://doi.org/10.1038/s41586-023-05791-5>
- Jayakodi M, Lu Q, Pidón H, Rabanus-Wallace MT, Bayer M, Lux T, Guo Y, Jaegle B, Badea A, Bekele W, Brar GS, Braune K, Bunk B, Chalmers KJ, Chapman B, Jørgensen ME, Feng J-W, Feser M, Fiebig A, Gundlach H, Guo W, Haberer G, Hansson M, Himmelbach A, Höffie I, Höffie RE, Hu H, Isobe S, König P, Kale SM, Kamal N, Keeble-Gagnère G, Keller B, Knauff M, Koppolu R, Krattinger SG, Kumlehn J, Langridge P, Li C, Marone MP, Maurer A, Mayer KFX, Melzer M, Muehlbauer GJ, Murozuka E, Padmarasu S, Perovic D, Pillen K, Pin PA, Pozniak CJ, Ramsay L, Pidas PR, Rutten T, Sakuma S, Sato K, Schüler D, Schmutzer T, Scholz U, Schreiber M, Shirasawa K, Simpson C, Skadhauge B, Spannagl M, Steffenson BJ, Thomsen HC, Tibbitts JF, Nielsen MTS, Trautewig C, Vequaud D, Voss C, Wang P, Waugh R, Westcott S, Rasmussen MW, Zhang R, Zhang X-Q, Wicker T, Dockter C, Mascher M, Stein N (2024) Structural variation in the pangenome of wild and domesticated barley. *Nature* 636:654–662. <https://doi.org/10.1038/s41586-024-08187-1>
- Jeffares DC, Jolly C, Hoti M, Speed D, Shaw L, Rallies C, Balloux F, Dessimoz C, Bähler J, Sedlazeck FJ (2017) Transient structural variations have strong effects on quantitative traits and reproductive isolation in fission yeast. *Nat Commun* 8:14061–14061. <https://doi.org/10.1038/ncomms14061>
- Jensen SE, Charles JR, Muleta K, Bradbury PJ, Casstevens T, Deshpande SP, Gore MA, Gupta R, Ilut DC, Johnson L, Lozano R, Miller Z, Ramu P, Rathore A, Romay MC, Upadhyaya HD, Varshney RK, Morris GP, Pressoir G, Buckler ES, Ramstein GP (2020) A sorghum practical haplotype graph facilitates genome-wide imputation and cost-effective genomic prediction. *Plant Genome* 13:e20009. <https://doi.org/10.1002/tpg2.20009>
- Jiao C, Xie X, Hao C, Chen L, Xie Y, Garg V, Zhao L, Wang Z, Zhang Y, Li T, Fu J, Chitkineni A, Hou J, Liu H, Dwivedi G, Liu X, Jia J, Mao L, Wang X, Appels R, Varshney RK, Guo W, Zhang X (2025) Pan-genome bridges wheat structural variations with habitat and breeding. *Nature* 637:384–393. <https://doi.org/10.1038/s41586-024-08277-0>
- Kaul S, Koo HL, Jenkins J, Rizzo M, Rooney T, Tallon LJ, Feldblyum T, Nierman W, Benito MI, Lin X, Town CD, Venter JC, Fraser CM, Tabata S, Nakamura Y, Kaneko T, Sato S, Asamizu E, Kato T, Kotani H, Sasamoto S, Ecker JR, Theologis A, Federspiel NA, Palm CJ, Osborne BI, Shinn P, Dewar K, Kim CJ, Buehler E, Dunn P, Chao Q, Chen H, Vysotskaia VS, Lenz CA, Hansen NF, Liu SX, Alta H, Sakano H, Lam B, Pham PK, Nguyen M, Yu G, Chen H, Southwick A, Lee JM, Miranda M, Toriumi MJ, Davis RW, Palm CJ, Conway AB, Conn L, Hootan A, Wambutt R, Murphy G, Düsterhöft A, Stiekema W, Pohl T, Entian KD, Terryn N, Volckaert G, Salanoubat M, Choise N, Artiguenave F, Weissenbach J, Quetier F, Rieger M, Ansong W, Unsel M, Fartmann B, Valle G, Wilson RK, Sekhon M, Pepin K, Murray J, Johnson D, Hillier L, de la Bastide M, Huang E, Spiegel L, Gnoj L, Habermann K, Dedhia N, Parnell L, Preston R, Marra M, McCombie WR, Chen E, Martienssen R, Mayer K, Lemcke K (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana**. *Nature* 408:796–815. <https://doi.org/10.1038/35048692>
- Khan AW, Garg V, Roorkiwal M, Golicz AA, Edwards D, Varshney RK (2020) Super-pangenome by integrating the wild side of a species

- for accelerated crop improvement. *Trends Plant Sci* 25:148–158. <https://doi.org/10.1016/j.tplants.2019.10.012>
- Khan AW, Garg V, Sun S, Gupta S, Dudchenko O, Roorkiwal M, Chitkineni A, Bayer PE, Shi C, Upadhyaya HD, Bohra A, Bhara-dwaj C, Mir RR, Baruch K, Yang B, Coyne CJ, Bansal KC, Nguyen HT, Ronen G, Aiden EL, Veneklaas E, Siddique KHM, Liu X, Edwards D, Varshney RK (2024) Cicer super-pangenome provides insights into species evolution and agronomic trait loci for crop improvement in chickpea. *Nat Genet* 56:1225–1234. <https://doi.org/10.1038/s41588-024-01760-4>
- Khoury CK, Brush S, Costich DE, Curry HA, Haan S, Engels JMM, Guarino L, Hoban S, Mercer KL, Miller AJ, Nabhan GP, Perales HR, Richards C, Riggins C, Thormann I (2022) Crop genetic erosion: understanding and responding to loss of crop diversity. *New Phytol* 233:84–118. <https://doi.org/10.1111/nph.17733>
- King GJ, Amoah S, Kurup S (2010) Exploring and exploiting epigenetic variation in crops. *Genome* 53:856–868. <https://doi.org/10.1139/G10-059>
- Kopalli V, Arslan K, Morales-Díaz N, Zanini SF, Golicz AA. Towards a Standardized Framework for Pangenome Graph Evaluation: assessing crop plant pangenome variation graph construction from multiple assemblies. 2025. [Accepted]
- Lam HM, Xu X, Liu X et al (2010) Resequencing of 31 wild and cultivated soybean genomes identifies patterns of genetic diversity and selection. *Nat Genet* 42:1053–1059. <https://doi.org/10.1038/ng.715>
- Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, Funke R, Gage D, Harris K, Heaford A, Howland J, Kann L, Lehoczky J, LeVine R, McEwan P, McKernan K, Meldrim J, Mesirov JP, Miranda C, Morris W, Naylor J, Raymond C, Rosetti M, Santos R, Sheridan A, Sougnez C, Stange-Thomann N, Stojanovic N, Subramanian A, Wyman D, Rogers J, Sulston J, Ainscough R, Beck S, Bentley D, Burton J, Clee C, Carter N, Coulson A, Deadman R, Deloukas P, Dunham A, Dunham I, Durbin R, French L, Grafham D, Gregory S, Hubbard T, Humphray S, Hunt A, Jones M, Lloyd C, McMurray A, Matthews L, Mercer S, Milne S, Mullikin JC, Mungall A, Plumb R, Ross M, Shownkeen R, Sims S, Waterston RH, Wilson RK, Hillier LW, McPherson JD, Marra MA, Mardis ER, Fulton LA, Chinwalla AT, Pepin KH, Gish WR, Chissole SL, Wendl MC, Delehaunty KD, Miner TL, Delehaunty A, Kramer JB, Cook LL, Fulton RS, Johnson DL, Minx PJ, Clifton SW, Hawkins T, Branscomb E, Predki P, Richardson P, Wenning S, Slezak T, Doggett N, Cheng J-F, Olsen A, Lucas S, Elkin C, Uberbacher E, Frazier M (2001) Initial sequencing and analysis of the human genome. *Nature* 409:860–921. <https://doi.org/10.1038/35057062>
- Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9:357–359. <https://doi.org/10.1038/nmeth.1923>
- Leonard AS, Crysnanto D, Mapel XM, Bhati M, Pausch H (2023) Graph construction method impacts variation representation and analyses in a bovine super-pangenome. *Genome Biol*. <https://doi.org/10.1186/s13059-023-02969-y>
- Liao W-W, Ebler J, Doerr D, Haukness M, Hickey G, Lu S, Lucas JK, Monlong J, Abel HJ, Buonaiuto S, Cheng H, Chu J, Colonna V, Eizenga JM, Feng X, Fulton RS, Garg S, Groza C, Guaracino A, Harvey WT, Heumos S, Howe K, Jain M, Lu T-Y, Martin FJ, Mitchell MW, Munson KM, Mwaniki MN, Novak AM, Olsen HE, Sirén J, Tomlinson C, Villani F, Vollger MR, Antonacci-Fulton LL, Baid G, Baker CA, Belyaeva A, Billis K, Carroll A, Chang P-C, Cook DE, Cook-Deegan RM, Diekhans M, Ebert P, Fedrigo O, Frankish A, Gao Y, Garrison NA, Giron CG, Green RE, Haggerty L, Hoekzema K, Hourlier T, Ji HP, Koenig BA, Korbel JO, Kordosky J, Koren S, Lee H, Magalhães H, Marco-Sola S, Marijon P, McCartney A, McDaniel J, Mountcastle J, Nattestad M, Nurk S, Olson ND, Popejoy AB, Puiu D, Rautiainen M, Regier AA, Rhie A, Sacco S, Sanders AD, Schneider VA, Schultz BI, Shafin K, Sofia HJ, Abou Tayoun AN, Thibaud-Nissen F, Tricomi FF, Wagner J, Wood JMD, Zimin AV, Bourque G, Chaisson MJP, Flicek P, Phillippy AM, Zook JM, Eichler EE, Haussler D, Wang T, Miga KH, Garrison E, Marshall T, Hall IM, Li H, Paten B (2023) A draft human pangenome reference. *Nature* 617:312–324. <https://doi.org/10.1038/s41586-023-05896-x>
- Li H (2018) Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 34:3094–3100. <https://doi.org/10.1093/bioinformatics/bty191>
- Li Y, Zhou G, Ma J, Jiang W, Jin L, Zhang Z, Guo Y, Zhang J, Sui Y, Zheng L, Zhang S, Zuo Q, Shi X, Li Y, Zhang W, Hu Y, Kong G, Hong H, Tan B, Song J, Liu Z, Wang Y, Ruan H, Yeung CKL, Liu J, Wang H, Li Z, Guan R, Wang K, Li WC, Chang R, Jiang Z, Jackson SA, Li R, Qiu L (2014) De novo assembly of soybean wild relatives for pan-genome analysis of diversity and agronomic traits. *Nat Biotechnol* 32:1045–1052. <https://doi.org/10.1038/nbt.2979>
- Li P, Quan X, Jia G, Xiao J, Cloutier S, You FM (2016) RGAugury: a pipeline for genome-wide prediction of resistance gene analogs (RGAs) in plants. *BMC Genomics* 17:852–852. <https://doi.org/10.1186/s12864-016-3197-x>
- Li H, Feng X, Chu C (2020) The design and construction of reference pangenome graphs with minigraph. *Genome Biol* 21:265–265. <https://doi.org/10.1186/s13059-020-02168-z>
- Li H, Wang S, Chai S, Yang Z, Zhang Q, Xin H, Xu Y, Lin S, Chen X, Yao Z, Yang Q, Fei Z, Huang S, Zhang Z (2022) Graph-based pan-genome reveals structural and sequence variations related to agronomic traits and domestication in cucumber. *Nat Commun* 13:682–682. <https://doi.org/10.1038/s41467-022-28362-0>
- Li N, He Q, Wang J, Wang B, Zhao J, Huang S, Yang T, Tang Y, Yang S, Aisimutuola P, Xu R, Hu J, Jia C, Ma K, Li Z, Jiang F, Gao J, Lan H, Zhou Y, Zhang X, Huang S, Fei Z, Wang H, Li H, Yu Q (2023a) Super-pangenome analyses highlight genomic diversity and structural variation across wild and cultivated tomato species. *Nat Genet* 55:852–860. <https://doi.org/10.1038/s41588-023-01340-y>
- Li X, Liu X, Fan Y, Li S, Yu M, Qian M, Chen Y, Chen H, Li X, Liu B, Xu X, Qu C, Li J, Lu K (2023b) Development of a target capture sequencing SNP genotyping platform for genetic analysis and genomic breeding in rapeseed. *Crop J* 11:499–510. <https://doi.org/10.1016/j.cj.2022.08.008>
- Li H (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv preprint*. <https://doi.org/10.48550/arxiv.1303.3997>
- Lieberman-Lazarovich M, Kaiserli E, Bucher E, Mladenov V (2022) Natural and induced epigenetic variation for crop improvement. *Curr Opin Plant Biol* 70:102297–102297. <https://doi.org/10.1016/j.pbi.2022.102297>
- Liu Y, Du H, Li P, Shen Y, Peng H, Liu S, Zhou GA, Zhang H, Liu Z, Shi M, Huang X, Li Y, Zhang M, Wang Z, Zhu B, Han B, Liang C, Tian Z (2020) Pan-genome of wild and cultivated soybeans. *Cell*. <https://doi.org/10.1016/j.cell.2020.05.023>
- Liu M, Zhang F, Lu H, Xue H, Dong X, Li Z, Xu J, Wang W, Wei C (2024a) PPanG: a precision pangenome browser enabling nucleotide-level analysis of genomic variations in individual genomes and their graph-based pangenome. *BMC Genomics* 25:405. <https://doi.org/10.1186/s12864-024-10302-5>
- Liu Z, Wang N, Su Y, Long Q, Peng Y, Shangguan L, Zhang F, Cao S, Wang X, Ge M, Xue H, Ma Z, Liu W, Xu X, Li C, Cao X, Ahmad B, Su X, Liu Y, Huang G, Du M, Liu Z, Gan Y, Sun L, Fan X, Zhang C, Zhong H, Leng X, Ren Y, Dong T, Pei D, Wu X, Jin Z,

- Wang Y, Liu C, Chen J, Gaut B, Huang S, Fang J, Xiao H, Zhou Y (2024b) Grapevine pangenome facilitates trait genetics and genomic breeding. *Nat Genet* 56:2804–2814. <https://doi.org/10.1038/s41588-024-01967-5>
- Liu D, Luo C, Dai R, Huang X, Chen X, He L, Mao H, Li J, Zhang L, Yang QY, Mei Z (2025) AMIR: a multi-omics data platform for *Asteraceae* plants genetics and breeding research. *Nucleic Acids Res* 53:D1563–D1575. <https://doi.org/10.1093/nar/gkae833>
- Long W, He Q, Wang Y, Wang J, Yuan Z, Wang M, Chen W, Luo L, Luo L, Xu W, Li Y, Li W, Yan L, Cai Y, Du H, Xie H (2024) Genome evolution and diversity of wild and cultivated rice species. *Nat Commun* 15:9994–9916. <https://doi.org/10.1038/s41467-024-54427-3>
- Lu K, Wei L, Li X, Wang Y, Wu J, Liu M, Zhang C, Chen Z, Xiao Z, Jian H, Cheng F, Zhang K, Du H, Cheng X, Qu C, Qian W, Liu L, Wang R, Zou Q, Ying J, Xu X, Mei J, Liang Y, Chai Y-R, Tang Z, Wan H, Ni Y, He Y, Lin N, Fan Y, Sun W, Li N-N, Zhou G, Zheng H, Wang X, Paterson AH, Li J (2019) Whole-genome resequencing reveals *Brassica napus* origin and genetic loci involved in its improvement. *Nat Commun* 10:1154. <https://doi.org/10.1038/s41467-019-09134-9>
- Lu Y, Wang J, Chen B, Mo S, Lian L, Luo Y, Ding D, Ding Y, Cao Q, Li Y, Li Y, Liu G, Hou Q, Cheng T, Wei J, Zhang Y, Chen G, Song C, Hu Q, Sun S, Fan G, Wang Y, Liu Z, Song B, Zhu J-K, Li H, Jiang L (2021) A donor-DNA-free CRISPR/Cas-based approach to gene knock-up in rice. *Nat Plants* 7:1445–1452. <https://doi.org/10.1038/s41477-021-01019-4>
- MacNish TR, Al-Mamun HA, Bayer PE, McPhan C, Fernandez CGT, Upadhyaya SR, Liu S, Batley J, Parkin IAP, Sharpe AG, Edwards D (2025) *Brassica* Panache: a multi-species graph pangenome representing presence absence variation across forty-one *Brassica* genomes. *Plant Genome* 18:e20535. <https://doi.org/10.1002/tpg2.20535>
- McHale LK, Haun WJ, Xu WW, Bhaskar PB, Anderson JE, Hyten DL, Gerhardt DL, Jeddalo JA, Stupar RM (2012) Structural variants in the soybean genome localize to clusters of biotic stress-response genes. *Plant Physiol* 159:41295–41308. <https://doi.org/10.1104/pp.112.194605>
- Mohamedikbal S, Al-Mamun HA, Marsh JJ, Upadhyaya S, Danilevicz MF, Nguyen HT, Valliyodan B, Mahan A, Batley J, Edwards D (2024) Local haplotyping reveals insights into the genetic control of flowering time variation in wild and domesticated soybean. *Plant Genome* 17:e20528. <https://doi.org/10.1002/tpg2.20528>
- Montenegro JD, Golitz AA, Bayer PE, Hurgobin B, Lee H, Chan CK, Visendi P, Lai K, Doležel J, Batley J, Edwards D (2017) The pangenome of hexaploid bread wheat. *Plant J* 90:1007–1013. <https://doi.org/10.1111/tpj.13515>
- Morales-Díaz N, Sushko S, Campos-Dominguez L, Kopalli V, Golitz AA, Castanera R, Casacuberta JM (2025) Tandem LTR-retrotransposon structures are common and highly polymorphic in plant genomes. *Mob DNA* 16:10. <https://doi.org/10.1186/s13100-025-00347-y>
- Morgante M, Brunner S, Pea G, Fengler K, Zuccolo A, Rafalski A (2005) Gene duplication and exon shuffling by helitron-like transposons generate intraspecies diversity in maize. *Nat Genet* 37:997–1002. <https://doi.org/10.1038/ng1615>
- Noshay JM, Springer NM (2021) Stories that can't be told by SNPs; DNA methylation variation in plant populations. *Curr Opin Plant Biol* 61:101989–101989. <https://doi.org/10.1016/j.pbi.2020.101989>
- Noshay JM, Liang Z, Zhou P, Crisp PA, Marand AP, Hirsch CN, Schmitz RJ, Springer NM (2021) Stability of DNA methylation and chromatin accessibility in structurally diverse maize genomes. *G3*. <https://doi.org/10.1093/g3journal/jkab190>
- Orantes-Bonilla M, Makhoul M, Lee H, Chawla HS, Vollrath P, Langstroff A, Sedlazeck FJ, Zou J, Snowdon RJ (2022) Frequent spontaneous structural rearrangements promote rapid genome diversification in a *Brassica napus* F1 generation. *Front Plant Sci* 13:1057953. <https://doi.org/10.3389/fpls.2022.1057953>
- Peng Y, Yan GL, Deng C, Wang C, Wang Y, Kang L, Zhou P, Yu K, Dong X, Liu X, Sun Z, Peng Y, Zhao J, Deng D, Xu Y, Li Y, Jiang Q, Li Y, Wei L, Wang J, Ma J, Hao M, Li W, Kang H, Peng Z, Liu D, Jia J, Zheng Y, Ma T, Wei Y, Lu F, Ren C (2022) Reference genome assemblies reveal the origin and evolution of allohexaploid oat. *Nat Genet* 54:1248–1258. <https://doi.org/10.1038/s41588-022-01127-7>
- Perez-Fons L, Ovalle TM, Drapal M, Ospina MA, Gkanogiannis A, Bohorquez-Chaux A, Becerra Lopez-Lavalle LA, Fraser PD (2023) Integrated genetic and metabolic characterization of Latin American cassava (*Manihot esculenta*) germplasm. *Plant Physiol* 192:2672–2686. <https://doi.org/10.1093/plphys/kiad269>
- Poplin R, Chang P-C, Alexander D, Schwartz S, Colthurst T, Ku A, Newburger D, Djimco J, Nguyen N, Afshar PT, Gross SS, Dorfman L, McLean CY, DePristo MA (2018) A universal SNP and small-indel variant caller using deep neural networks. *Nat Biotechnol* 36:983–987. <https://doi.org/10.1038/nbt.4235>
- Qian Y, Hu W, Liao J, Zhang J, Ren Q (2019) The dynamics of DNA methylation in the maize (*Zea mays* L.) inbred line B73 response to heat stress at the seedling stage. *Biochem Biophys Res Commun* 512:742–749. <https://doi.org/10.1016/j.bbrc.2019.03.150>
- Qin P, Lu H, Du H, Wang H, Chen W, Chen Z, He Q, Ou S, Zhang H, Li X, Li X, Li Y, Liao Y, Gao Q, Tu B, Yuan H, Ma B, Wang Y, Qian Y, Fan S, Li W, Wang J, He M, Yin J, Li T, Jiang N, Chen X, Chengzhi L, Li S (2021) Pan-genome analysis of 33 genetically diverse rice accessions reveals hidden genomic variations. *Cell* 184:3542–3558.e3516. <https://doi.org/10.1016/j.cell.2021.04.046>
- Rabiner LR (1989) A tutorial on hidden Markov models and selected applications in speech recognition. *Proc IEEE* 77:257–286. <https://doi.org/10.1109/5.18626>
- Rijzaani H, Bayer PE, Rouard M, Doležel J, Batley J, Edwards D (2022) The pangenome of banana highlights differences between genera and genomes. *Plant Genome* 15:e20100. <https://doi.org/10.1002/tpg2.20100>
- Samans B, Chalhoub B, Snowdon RJ (2017) Surviving a genome collision: genomic signatures of allopolyploidization in the recent crop species *Brassica napus*. *Plant Genome* 10:1–15. <https://doi.org/10.3835/plantgenome2017.02.0013>
- Sarashetti P, Lipovac J, Tomas F, Šikić M, Liu J (2024) Evaluating data requirements for high-quality haplotype-resolved genomes for creating robust pangenome references. *Genome Biol* 25:312. <https://doi.org/10.1186/s13059-024-03452-y>
- Saxena RK, Edwards D, Varshney RK (2014) Structural variations in plant genomes. *Brief Funct Genomics* 13:296–307. <https://doi.org/10.1093/bfpg/elu016>
- Secomandi S, Gallo GR, Rossi R, Rodríguez Fernandes C, Jarvis ED, Bonisoli-Alquati A, Gianfranceschi L, Formenti G (2025) Pangenome graphs and their applications in biodiversity genomics. *Nat Genet* 57:147719. <https://doi.org/10.1038/s41588-024-02029-6>
- Sentieon (2018) Sentieon DNaseq <https://www.sentieon.com/products/>. Accessed 15 May 2025.
- Shang L, Li X, He H, Yuan Q, Song Y, Wei Z, Lin H, Hu M, Zhao F, Zhang C, Li Y, Gao H, Wang T, Liu X, Zhang H, Zhang Y, Cao S, Yu X, Zhang B, Zhang Y, Tan Y, Qin M, Ai C, Yang Y, Zhang B, Hu Z, Wang H, Lv Y, Wang Y, Ma J, Wang Q, Lu H, Wu Z, Liu S, Sun Z, Zhang H, Guo L, Li Z, Zhou Y, Li J, Zhu Z, Xiong G, Ruan J, Qian Q (2022) A super pan-genomic landscape of rice. *Cell Res* 32:878–896. <https://doi.org/10.1038/s41422-022-00685-z>

- She H, Liu Z, Xu Z, Zhang H, Wu J, Cheng F, Wang X, Qian W (2024) Pan-genome analysis of 13 *Spinacia* accessions reveals structural variations associated with sex chromosome evolution and domestication traits in spinach. *Plant Biotechnol J* 22:3102–3117. <https://doi.org/10.1111/pbi.14433>
- Sheikhizadeh Anari S, de Ridder D, Schranz ME, Smit S (2018) Efficient inference of homologs in large eukaryotic pan-proteomes. *BMC Bioinformatics* 19:340–340. <https://doi.org/10.1186/s12859-018-2362-4>
- Sibbesen JA, Eizenga JM, Novak AM, Sirén J, Chang X, Garrison E, Paten B (2023) Haplotype-aware pantranscriptome analyses using spliced pangenome graphs. *Nat Methods*. <https://doi.org/10.1038/s41592-022-01731-9>
- Song J-M, Guan Z, Hu J, Guo C, Yang Z, Wang S, Liu D, Wang B, Lu S, Zhou R, Xie W-Z, Cheng Y, Zhang Y, Liu K, Yang Q-Y, Chen L-L, Guo L (2020) Eight high-quality genomes reveal pan-genome architecture and ecotype differentiation of *Brassica napus*. *Nat Plants* 6:34–45. <https://doi.org/10.1038/s41477-019-0577-7>
- Songsomboon K, Brenton Z, Heuser J, Kresovich S, Shakoor N, Mockler T, Cooper EA (2021) Genomic patterns of structural variation among diverse genotypes of *Sorghum bicolor* and a potential role for deletions in local adaptation. *G3 Bethesda*. <https://doi.org/10.1093/g3journal/jkab154>
- Springer NM, Ying K, Fu Y, Ji T, Yeh C-T, Jia Y, Wu W, Richmond T, Kitzman J, Rosenbaum H, Iniguez AL, Barbazuk WB, Jeddeloh JA, Nettleton D, Schnable PS (2009) Maize inbreds exhibit high levels of copy number variation (CNV) and presence/absence variation (PAV) in genome content. *PLoS Genet* 5:e1000734–e1000734. <https://doi.org/10.1371/journal.pgen.1000734>
- Steuernagel B, Witek K, Krattinger SG, Ramirez-Gonzalez RH, Schoonbeek H, Yu G, Bagges E, Witek AI, Yadav I, Krasileva KV, Jones JDG, Uauy C, Keller B, Ridout CJ, Wulff BBH (2020) The NLR-annotator tool enables annotation of the intracellular immune receptor repertoire. *Plant Physiol* 183:468–482. <https://doi.org/10.1104/pp.19.01273>
- Su Y, Yang X, Wang Y, Li J, Long Q, Cao S, Wang X, Liu Z, Huang S, Chen Z, Peng Y, Zhang F, Xue H, Cao X, Zhang M, Yisilam G, Chu Z, Gao Y, Zhou Y, Liu Z, Xiao H, Tian X (2024) Phased telomere-to-telomere reference genome and pangenome reveal an expansion of resistance genes during apple domestication. *Plant Physiol* 195:2799–2814. <https://doi.org/10.1093/plphys/kiac258>
- Sun J, Wang Y, Guo C, Ge R, Naren T, Jiang L (2024) CRISPR/Cas9 effectively generate chromosome structural variations in rice protoplasts. *Crop Sci* 64:3430–3437. <https://doi.org/10.1002/csc2.21334>
- Tang D, Jia Y, Zhang J, Li H, Cheng L, Wang P, Bao Z, Liu Z, Feng S, Zhu X, Li D, Zhu G, Wang H, Zhou Y, Zhou Y, Bryan GJ, Buell CR, Zhang C, Huang S (2022) Genome evolution and diversity of wild and cultivated potatoes. *Nature* 606:535–541. <https://doi.org/10.1038/s41586-022-04822-x>
- Tao Y, Luo H, Xu J, Cruickshank A, Zhao X, Teng F, Hathorn A, Wu X, Liu Y, Shatte T, Jordan D, Jing H, Mace E (2021) Extensive variation within the pan-genome of cultivated and wild sorghum. *Nat Plants* 7:766–773. <https://doi.org/10.1038/s41477-021-00925-x>
- Tariq A, Meng M, Jiang X, Bolger A, Beier S, Buchmann JP, Fernie AR, Wen W, Usadel B (2024) In-depth exploration of the genomic diversity in tea varieties based on a newly constructed pangenome of *Camellia sinensis*. *Plant J* 119:2096–2115. <https://doi.org/10.1111/tbj.16874>
- Tay Fernandez CG, Nestor BJ, Danilevicz MF, Marsh JJ, Petereit J, Bayer PE, Batley J, Edwards D (2022) Expanding gene-editing potential in crop improvement with pangenomes. *Int J Mol Sci* 23:2276. <https://doi.org/10.3390/ijms23042276>
- Tettelin H, Masignani V, Cieslewicz MJ, Donati C, Medini D, Ward NL, Angiuoli SV, Crabtree J, Jones AL, Durkin AS, DeBoy RT, Daviden TM, Mora M, Scarselli M, MyR I, Peterson JD, Hauser CR, Sundaram JP, Nelson WC, Madupu R, Brinkac LM, Dodson RJ, Rosovitz MJ, Sullivan SA, Daugherty SC, Haft DH, Selengut J, Gwinn ML, Zhou L, Zafar N, Khouri H, Radune D, Dimitrov G, Watkins K, O'Connor KJB, Smith S, Utterback TR, White O, Rubens CE, Grandi G, Madoff LC, Kasper DL, Telford JL, Wessels MR, Rappuoli R, Fraser CM (2005) Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: Implications for the microbial ‘pan-genome.’ *Proc Natl Acad Sci* 102:13950–13955. <https://doi.org/10.1073/pnas.0506758102>
- Tørresen OK, Star B, Mier P, Andrade-Navarro MA, Bateman A, Jarrot P, Gruca A, Grynberg M, Kajava AV, Promponas VJ, Anisimova M, Jakobsen KS, Linke D (2019) Tandem repeats lead to sequence assembly errors and impose multi-level challenges for genome and protein databases. *Nucleic Acids Res* 47:10994–11006. <https://doi.org/10.1093/nar/gkz841>
- Vafadarshamasbi U, Mace E, Jordan D, Crisp PA (2022) Decoding the sorghum methylome: understanding epigenetic contributions to agronomic traits. *Biochem Soc Trans* 50:583–596. <https://doi.org/10.1042/BST20210908>
- van der Peer Y, Ashman T-L, Soltis PS, Soltis DE (2021) Polyploidy: an evolutionary and ecological force in stressful times. *Plant Cell* 33:11–26. <https://doi.org/10.1093/plcell/koaa015>
- van Workum D-JM, Mehrem SL, Snoek BL, Alderkamp MC, Lapin D, Mulder FFM, Van den Ackerveken G, de Ridder D, Schranz ME, Smit S (2024) *Lactuca* super-pangenome reduces bias towards reference genes in lettuce research. *BMC Plant Biol* 24:1019–1019. <https://doi.org/10.1186/s12870-024-05712-2>
- Varshney RK, Roorkiwal M, Sun S, Bajaj P, Chitkineni A, Thudi M, Singh NP, Du X, Upadhyaya HD, Khan AW, Wang Y, Garg V, Fan G, Cowling WA, Crossa J, Gentzbittel L, Voss-Fels KP, Valluri VK, Sinha P, Singh VK, Ben C, Rathore A, Punna R, Singh MK, Tar'an B, Bharadwaj C, Yasin M, Pithia MS, Singh S, Soren KR, Kudapa H, Jarquin D, Cubry P, Hickey LT, Dixit GP, Thuillet AC, Hamwieh A, Kumar S, Deokar AA, Chaturvedi SK, Francis A, Howard R, Chattopadhyay D, Edwards D, Lyons E, Vigouroux Y, Hayes BJ, von Wettberg E, Datta SK, Yang H, Nguyen HT, Wang J, Siddique KHM, Mohapatra T, Benetzen JL, Xu X, Liu X (2021) A chickpea genetic variation map based on the sequencing of 3366 genomes. *Nature* 599:622–627. <https://doi.org/10.1038/s41586-021-04066-1>
- Vollrath P, Chawla HS, Schiessl SV, Gabur I, Lee HT, Snowdon RJ, Obermeier C (2021a) A novel deletion in FLOWERING LOCUS T modulates flowering time in winter oilseed rape. *Theor Appl Genet* 134:1217–1231. <https://doi.org/10.1007/s00122-021-03768-4>
- Vollrath P, Chawla HS, Alnajjar D, Gabur I, Lee HT, Weber S, Ehrig L, Koopmann B, Snowdon RJ, Obermeier C (2021b) Dissection of quantitative blackleg resistance reveals novel variants of resistance gene Rlm9 in elite *Brassica napus*. *Front Plant Sci*. <https://doi.org/10.3389/fpls.2021.749491>
- Wang W, Mauleon R, Hu Z, Chebotarov D, Tai S, Wu Z et al (2018) Genomic variation in 3010 diverse accessions of Asian cultivated rice. *Nature* 557:43–49. <https://doi.org/10.1038/s41586-018-0063-9>
- Wang Q, Xu J, Pu X, Lv H, Liu Y, Ma H, Wu F, Wang Q, Feng X, Liu T, Tang Q, Liu Y, Lu Y (2021) Maize DNA methylation in response to drought stress is involved in target gene expression and alternative splicing. *Int J Mol Sci* 22:8285. <https://doi.org/10.3390/ijms22158285>
- Wang S, Qian Y-Q, Zhao R-P, Chen L-L, Song J-M (2023) Graph-based pan-genomes: increased opportunities in plant genomics. *J Exp Bot* 74:24–39. <https://doi.org/10.1093/jxb/erac412>

- Wang Y, Li P, Zhu Y, Zhang F, Zhang S, He Y, Wu Y, Lin Y, Wang H, Ren W, Wang L, Yang Y, Wang R, Zheng P, Liu Y, Wang S, Yue J (2024) Graph-based pangenome of *Actinidia chinensis* reveals structural variations mediating fruit degreening. *Adv Sci* 11:e2400322. <https://doi.org/10.1002/advs.202400322>
- Wick RR, Schultz MB, Zobel J, Holt KE (2015) Bandage: interactive visualization of de novo genome assemblies. *Bioinformatics* 31:3350–3352. <https://doi.org/10.1093/bioinformatics/btv383>
- Wu S, Sun H, Gao L, Branham S, McGregor C, Renner SSS, Xu Y, Kousik C, Wechter WP, Levi A, Fei Z (2023) A *Citrullus* genus super-pangenome reveals extensive variations in wild and cultivated watermelons and sheds light on watermelon evolution and domestication. *Plant Biotechnol J* 21:1926–1928. <https://doi.org/10.1111/pbi.14120>
- Xu J, Chen G, Hermanson PJ, Xu Q, Sun C, Chen W, Kan Q, Li M, Crisp PA, Yan J, Li L, Springer NM, Li Q (2019) Population-level analysis reveals the widespread occurrence and phenotypic consequence of DNA methylation variation not tagged by genetic variation in maize. *Genome Biol.* <https://doi.org/10.1186/s13059-019-1859-0>
- Xu G, Lyu J, Li Q, Liu H, Wang D, Zhang M, Springer NM, Ross-Ibarra J, Yang J (2020) Evolutionary and functional genomics of DNA methylation in maize domestication and improvement. *Nat Commun.* <https://doi.org/10.1038/s41467-020-19333-4>
- Yan H, Jin Y, Yu H, Wang C, Wu B, Jones CS, Wang X, Xie Z, Huang L (2024) Genomic selection for agronomical phenotypes using genome-wide SNPs and SVs in pearl millet. *Theor Appl Genet* 137:244–244. <https://doi.org/10.1007/s00122-024-04754-2>
- Yao W, Li G, Zhao H, Wang G, Lian X, Xie W (2015) Exploring the rice dispensable genome using a metagenome-like assembly strategy. *Genome Biol* 16:187–187. <https://doi.org/10.1186/s13059-015-0757-3>
- Yao E, Blake VC, Cooper L, Wight CP, Michel S, Cagirici HB, Lazo GR, Birkett CL, Waring DJ, Jannink J-L, Holmes I, Waters AJ, Eickholt DP, Sen TZ (2022) GrainGenes: a data-rich repository for small grains genetics and genomics. *Database.* <https://doi.org/10.1093/database/baac034>
- Yildiz G, Zanini SF, Weber S, Kopalli V, Kox T, Abbadi A, Snowdon RJ, Golicz AA (2025) Graphical pangenomics-enabled characterization of structural variant impact on gene expression in *Brassica napus*. *Theor Appl Genet* 138:91. <https://doi.org/10.1007/s00122-025-04867-2>
- Yuan Y, Ma RK-K, Chan T-F (2023) PanGraphViewer: a versatile tool to visualize pangenome graphs. *bioRxiv.* <https://doi.org/10.1101/2023.03.30.534931>
- Zanini SF, Bayer PE, Wells R, Snowdon RJ, Batley J, Varshney RK, Nguyen HT, Edwards D, Golicz AA (2022) Pangenomics in crop improvement—from coding structural variations to finding regulatory variants with pangenome graphs. *Plant Genome* 15:e20177. <https://doi.org/10.1002/tpg2.20177>
- Zhang H, Lang Z, Zhu JK (2018) Dynamics and function of DNA methylation in plants. *Nat Rev Mol Cell Biol* 19:489–506. <https://doi.org/10.1038/s41580-018-0016-z>
- Zhang Y, Yang Z, He Y, Liu D, Liu Y, Liang C, Xie M, Jia Y, Ke Q, Zhou Y, Cheng X, Huang J, Liu L, Xiang Y, Raman H, Kliebenstein DJ, Liu S, Yang Q-Y (2024a) Structural variation reshapes population gene expression and trait variation in 2,105 *Brassica napus* accessions. *Nat Genet* 56:2538–2550. <https://doi.org/10.1038/s41588-024-01957-7>
- Zhang Y, Zhao M, Tan J, Huang M, Chu X, Li Y, Han X, Fang T, Tian Y, Jarret R, Lu D, Chen Y, Xue L, Li X, Qin G, Li B, Sun Y, Deng XW, Deng Y, Zhang X, He H (2024b) Telomere-to-telomere *Citrullus* super-pangenome provides direction for watermelon breeding. *Nat Genet* 56:1750–1761. <https://doi.org/10.1038/s41588-024-01823-6>
- Zhang Z, Gomes Viana JP, Zhang B, Walden KKO, Müller Paul H, Moose SP, Morris GP, Daum C, Barry KW, Shakoor N, Hudson ME (2024c) Major impacts of widespread structural variation on sorghum. *Genome Res* 20:286–299. <https://doi.org/10.1101/gr.278396.123>
- Zhang Z, Liu D, Li B, Wang W, Zhang J, Xin M, Hu Z, Liu J, Du J, Peng H, Hao C, Zhang X, Ni Z, Sun Q, Guo W, Yao Y (2024d) 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* 17:1038–1053. <https://doi.org/10.1016/j.molp.2024.05.006>
- Zhao T, Guan X, Hu Y, Zhang Z, Yang H, Shi X, Han J, Mei H, Wang L, Shao L, Wu H, Chen Q, Zhao Y, Pan J, Hao Y, Dong Z, Long X, Deng Q, Zhao S, Zhang M, Zhu Y, Ma X, Chen Z, Deng Y, Si Z, Li X, Zhang T, Gu F, Gu X, Fang L (2024) Population-wide DNA methylation polymorphisms at single-nucleotide resolution in 207 cotton accessions reveal epigenomic contributions to complex traits. *Cell Res* 34:859–872. <https://doi.org/10.1038/s41422-024-01027-x>
- Zheng L-Y, Guo X-S, He B, Sun L-J, Peng Y, Dong S-S, Liu T-F, Jiang S, Ramachandran S, Liu C-M, Jing H-C (2011) Genome-wide patterns of genetic variation in sweet and grain sorghum (*Sorghum bicolor*). *Genome Biol* 12:R114. <https://doi.org/10.1186/gb-2011-12-11-r114>
- Zhi P, Chang C (2021) Exploiting epigenetic variations for crop disease resistance improvement. *Front Plant Sci* 12:692328. <https://doi.org/10.3389/fpls.2021.692328>
- Zhong Y, Luo Y, Sun J, Qin X, Gan P, Zhou Z, Qian Y, Zhao R, Zhao Z, Cai W, Luo J, Chen LL, Song JM (2024) Pan-transcriptomic analysis reveals alternative splicing control of cold tolerance in rice. *Plant Cell* 36:2117–2139. <https://doi.org/10.1093/plcell/koae039>
- Zhou Y, Zhang Z, Bao Z, Li H, Lyu Y, Zan Y, Wu Y, Cheng L, Fang Y, Wu K, Zhang J, Lyu H, Lin T, Gao Q, Saha S, Mueller L, Fei Z, Städler T, Xu S, Zhang Z, Speed D, Huang S (2022) Graph pangenome captures missing heritability and empowers tomato breeding. *Nature* 606:527–534. <https://doi.org/10.1038/s41586-022-04808-9>
- Zhu X, Yang R, Liang Q, Yu Y, Wang T, Meng L, Wang P, Wang S, Li X, Yang Q, Guo H, Sui Q, Wang Q, Du H, Chen Q, Liang Z, Wu X, Zeng Q, Huang B (2025) Graph-based pangenome provides insights into structural variations and genetic basis of metabolic traits in potato. *Mol Plant* 18:590–602. <https://doi.org/10.1016/j.molp.2025.01.017>
- Zhuang Y, Wang X, Li X, Hu J, Fan L, Landis JB, Cannon SB, Grimwood J, Schmutz J, Jackson SA, Doyle JJ, Zhang XS, Zhang D, Ma J (2022) Phylogenomics of the genus *Glycine* sheds light on polyploid evolution and life-strategy transition. *Nat Plants* 8:233–244. <https://doi.org/10.1038/s41477-022-01102-4>
- Zimin AV, Marçais G, Puiu D, Roberts M, Salzberg SL, Yorke JA (2013) The MaSuRCA genome assembler. *Bioinformatics* 29:2669–2677. <https://doi.org/10.1093/bioinformatics/btt476>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.