Evolutionary Journal of the Linnean Society, 2025, **4(1)**, kzaf010
https://doi.org/10.1093/evolinnean/kzaf010
Advance access publication 19 June 2025
Perspective

Perspective

Pangenomes provide new insights into polyploidy in plants

Shengchen Shan[1,*], Jonathan P. Spoelhof[1], Paul D. Blischak[2], Jacqueline Batley[3], Pamela S. Soltis[1,4,5], Douglas E. Soltis[1,4,5,6], David Edwards[3]

[1]Florida Museum of Natural History, University of Florida, Gainesville, FL 32611, United States
[2]Bayer Crop Science, Chesterfield, MO 63017, United States
[3]School of Biological Sciences and Centre for Applied Bioinformatics, University of Western Australia, Perth, WA 6009, Australia
[4]Genetics Institute, University of Florida, Gainesville, FL 32610, United States
[5]Biodiversity Institute, University of Florida, Gainesville, FL 32611, United States
[6]Department of Biology, University of Florida, Gainesville, FL 32611, United States

[*]Corresponding author. Florida Museum of Natural History, University of Florida, Gainesville, FL, United States. E-mail: shan158538@ufl.edu

ABSTRACT

Polyploidy, also known as whole-genome duplication (WGD), is a significant evolutionary force in green plants, especially angiosperms. The dynamic nature of polyploid genomes generates genetic diversity and drives the evolution of novel traits and adaptations. Pangenomics is emerging as a major frontier in plant genome research, with a rapidly growing number of pangenomes for individual species and associated analyses providing novel agronomic and evolutionary insights. Polyploid genome analysis can be confounded by intraspecific variation when relying on a single reference genome assembly. The use of pangenomes that better represent the genomic diversity of a species helps overcome this limitation. However, a major gap remains between the number of pangenomic studies in polyploid compared to diploid species, despite the widespread prevalence of WGD, limiting the potential of the pangenome framework for characterizing and understanding polyploid genomes. Furthermore, most polyploid pangenome studies have focused on domesticated crop species, and natural populations have rarely been examined. In addition to applications in crop improvement, pangenomes can provide insights into the ecological and evolutionary impact of polyploidy. Here, we summarize recent pangenome studies in polyploid plants and highlight promising topics for future research. We hope this article will encourage the growth of pangenomic studies in polyploid systems, particularly in natural populations.

Keywords: gene loss; genome evolution; natural populations; pangenome; polyploidy; structural variation

WHAT ARE PANGENOMES?

A pangenome is the collection of genome sequence assemblies from multiple individuals within a species (pan, derived from the Greek word *πᾶν*, meaning whole). Because intraspecific genetic diversity can be very high in plants, a single reference genome is inadequate to capture the genetic diversity of a species, which illustrates the necessity of generating pangenomes to overcome this limitation (Golicz *et al.* 2016a, 2020, Bayer *et al.* 2020, Lei *et al.* 2021, Shi *et al.* 2023, Schreiber *et al.* 2024). The conceptual framework of pangenomes was first proposed for studying microbes (Tettelin *et al.* 2005). Since the publication of the first intraspecific plant genome comparison in soybean (Li *et al.* 2014), many pangenome studies in plants have been published (reviewed in Bayer *et al.* 2020, Lei *et al.* 2021, Yuan *et al.* 2021a, Zanini *et al.* 2022, Shi *et al.* 2023, Song *et al.* 2023, Schreiber *et al.* 2024). Although most of the focus has been on diploid species, some studies have generated pangenomes for polyploid crops, including cotton (*Gossypium* spp.) (Jin *et al.* 2023, Wang N. *et al.* 2023), potato (*Solanum tuberosum*) (Hoopes *et al.* 2022, Bozan *et al.* 2023, Sun *et al.* 2025), *Brassica napus* (Hurgobin *et al.* 2018, Bayer *et al.* 2021), camelina (*Camelina sativa*; Bird *et al.* 2025), *Malus* spp. (Li *et al.* 2025), *Chenopodium* spp. (quinoa and its relatives; Jaggi *et al.* 2025), and hexaploid wheat (*Triticum aestivum*) (Montenegro *et al.* 2017, Bayer *et al.* 2022a, Jiao *et al.* 2025).

Pangenome studies help characterize the diversity of genomic variation within a species. Single nucleotide polymorphisms (SNPs) and small indels (<50 bp) can be captured by short-read DNA sequencing. However, larger structural variants (SVs) are best identified by comparing genome assemblies between different individuals. SVs comprise presence/absence variants (PAVs),

Received 26 November 2024; revised 07 May 2025; accepted 10 June 2025

copy number variants (CNVs), inversions, and translocations (reviewed in Yuan *et al.* 2021a) and can lead to the presence or absence of genes between individuals. While many early pangenome studies focused on the coding portions of the genome, advances in long-read DNA sequencing now permit a more comprehensive analysis of structural variation that also includes noncoding and repetitive DNA (reviewed in Lei *et al.* 2021, Zanini *et al.* 2022).

Several key features are often examined in pangenome studies. First, the size of the pangenome is a fundamental characteristic that has been shown to vary between domesticated individuals and their wild relatives (e.g. Bayer *et al.* 2021), and even between distinct domestication events within the same species (e.g. Cortinovis *et al.* 2024). Most analyses suggest that plant pangenomes are closed (i.e. it is possible to predict the total number of genes in a species) (Golicz *et al.* 2020). This contrasts with many bacterial pangenomes that are open (i.e. the pangenome continues to expand indefinitely as more individuals are included) due to extensive horizontal transfer of genes in bacteria (Golicz *et al.* 2020). Second, within a pangenome, genes are categorized as either core (present in all individuals) or variable (present only in certain individuals; also referred to as dispensable/shell/private/cloud/noncore genes) (Fig. 1A). In both diploids and polyploids, core genes are usually enriched in housekeeping functions, and variable genes are related to stress response and local adaptation. For example, in the pangenome of hexaploid wheat (*Triticum aestivum*), the variable genes are enriched in functions related to response to environmental stress and defence response (Montenegro *et al.* 2017). In diploid *Brassica oleracea*, the variable genes are enriched in functions including disease resistance and flowering time (Golicz *et al.* 2016b). Variable genes are enriched for functions related to defence and development in diploid *Brachypodium distachyon* (Gordon *et al.* 2017).

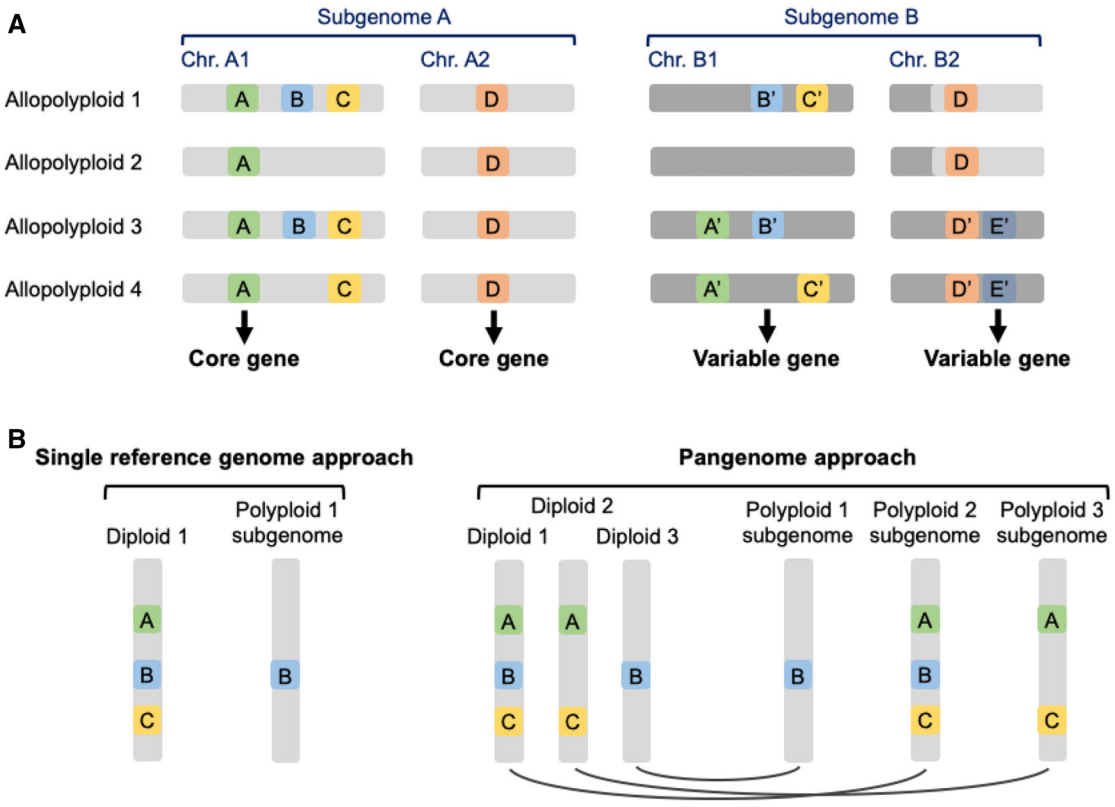

Figure 1. Pangenome analysis in diploids and polyploids. A, a pangenomic study of four allopolyploid individuals (1–4) reveals dynamic genomic changes. For each allopolyploid individual, two chromosomes (e.g. chr. A1 and chr. A2) from each subgenome (e.g. subgenome A) are shown. Each letter within the squares represents a gene. Homeologous genes are differentiated: for example, gene B on chr. A1 is homeologous to gene B′ on chr. B1. Core genes (e.g. gene A) are present in all individuals, whereas variable genes (e.g. gene B′ and gene E′) are present in only a subset of individuals. Nonreciprocal homeologous exchange (HE)—where a region on chr. B2 is replaced by the homeologous region on chr. A2, as seen in allopolyploid individuals 1 and 2—can lead to gene loss. In this example, gene E is absent on chr. A2 but present on chr. B2 (shown as E′). Nonreciprocal HE results in the loss of gene E′ in allopolyploids 1 and 2, making E′ a variable gene. In addition, one subgenome (e.g. subgenome B) often contains a higher proportion of variable genes than the other, a pattern observed in some polyploid systems. B, the pangenome approach allows for accurate characterization of genome evolutionary patterns following polyploidy. Using a single reference genome approach (e.g. diploid accession 1), genes A and C appear to be lost following polyploidy, a process referred to as fractionation (left panel). However, constructing pangenomes for both the diploid progenitors (accessions 1–3) and the polyploids (accessions 1–3) reveals that the gene presence/absence variation within the polyploid subgenomes can be explained by the interspecific variation of the diploid. This suggests that the genomic characteristics of one polyploid subgenome may be inherited from a specific diploid progenitor (inheritance indicated by connecting lines; right panel). In this scenario, the observed polyploid genomic features are more likely to be a result of parental legacy rather than genomic changes following polyploidy, a pattern that can only be revealed by using the pangenome approach.

Another important application of pangenomes is the use of gene PAV data for various downstream purposes, including phylogenetic analysis, genome-wide association studies (GWAS), and breeding (e.g. Hurgobin *et al.* 2018, Song *et al.* 2020, Li *et al.* 2021, Bozan *et al.* 2023). Pangenome-wide trait association studies have also captured greater heritability than single reference genomes (Edwards and Batley 2022, Jin *et al.* 2023, Cortinovis *et al.* 2024). Additionally, Zhou *et al.* (2022) highlighted the power of the graph pangenome (i.e. depicting sequence variations as branches in linear sequences) in tomato breeding and identified new genes contributing to soluble solid content, an important trait for both yield and flavour. Compared to using a single linear reference genome, employing the graph pangenome for GWAS increased the estimated heritability by 24% for gene expression and metabolite traits in tomato (Zhou *et al.* 2022). A pangenome study in wheat identified genes associated with environmental adaptation (*VRN-A1*) and grain hardness (*PIN*) (Jiao *et al.* 2025). Pangenomic analysis of alfalfa (*Medicago sativa*) revealed SVs associated with salt tolerance and quality traits (He *et al.* 2025).

Superpangenomes, which are collections of genomes from multiple phylogenetically closely related species, have been assembled for some crop species and their wild relatives to inform breeding efforts (Bayer *et al.* 2022b, Rijzaani *et al.* 2022, Bozan *et al.* 2023, Raza *et al.* 2023, Khan *et al.* 2024, Guo *et al.* 2025, Li *et al.* 2025). For example, Bozan *et al.* (2023) constructed a superpangenome of *Solanum* including 296 diverse potato accessions from 60 different *Solanum* species, representing both wild species and cultivars with varying ploidal levels (2*x*, 3*x*, 4*x*, and 5*x*). A superpangenome study of grapevine (*Vitis* spp.) identified downy mildew resistance genes, which will accelerate the breeding of this fruit crop (Guo *et al.* 2025).

Despite the broad application of pangenomics in crop improvement, there has been limited research on naturally occurring plants, where patterns of gene PAV often differ significantly from those observed in crops. *Arabidopsis thaliana*, a key model organism in functional biology, also serves as a powerful model in population genomics due to its broad geographical distribution. Multiple studies have explored the *Arabidopsis* pangenome (Jiao and Schneeberger 2020, Kang *et al.* 2023, Lian *et al.* 2024). For instance, Lian *et al.* (2024) examined the pangenome of *A. thaliana* using 69 accessions from Europe, Africa, and Asia, and found that 40% of gene families were variable. Importantly, even with a large sample size, the *Arabidopsis* pangenome remained open—the number of pangenome gene sets continued to grow as additional genomes were incorporated, without reaching a plateau even after all 69 genomes were added (Lian *et al.* 2024). This pattern contrasts with that found in crop species, which typically have closed pangenomes (e.g. Golicz *et al.* 2016b, 2020, Montenegro *et al.* 2017, Hurgobin *et al.* 2018, Bozan *et al.* 2023). Although Peng *et al.* (2022) observed an open pangenome in cottons (*Gossypium* spp.), this may be due to the relatively small sample size (eight genomes) and the inclusion of multiple species. Lian *et al.* (2024) revealed that the open pangenome in *Arabidopsis* is driven by a high percentage of genes (18%) unique to single accessions, and these genes may contribute to the adaptation of natural *Arabidopsis* populations to diverse environments.

Another example of a pangenomic study in natural populations is from *Amborella trichopoda*, the sole living sister species to all other extant flowering plants and native only to the remote island of New Caledonia. By analysing 10 individuals, Hu *et al.* (2022) revealed several unique features in the *Amborella* pangenome. Compared to crop species, *Amborella* has relatively few variable genes. Additionally, the variable genes in crops are frequently associated with both biotic and abiotic stress. In contrast, the variable genes of *Amborella* are mainly enriched in functions related to abiotic stress, with few variable genes associated with biotic stress (Hu *et al.* 2022). *Amborella* PAV genes associated with abiotic stress may play a role in environmental adaptation, and the relatively small and invariable set of disease resistance genes in *Amborella* may indicate limited pathogen pressure on the isolated island (Hu *et al.* 2022). In summary, pangenome studies in naturally occurring plants, such as *Arabidopsis* and *Amborella*, highlight the necessity of expanding pangenome research beyond crop species. Investigating pangenomes in natural populations provides a powerful tool for addressing key ecological and evolutionary questions.

BUILDING PANGENOMES

There are three general approaches for pangenome assembly (Table 1; reviewed by Bayer *et al.* 2020, Shi *et al.* 2023). In Box 1, we summarize currently used programs and packages for constructing and analysing pangenomes, aiming to offer a useful starting point for those new to the field. The first approach, used to estimate the gene content of a species, is the iterative assembly method. In this method, Illumina paired reads from multiple individuals of a species are mapped to a single reference genome, with unaligned reads being assembled using a metagenome-aware assembler into novel contigs. The pangenome is then constructed by annotating and adding these novel contigs to the original

Table 1. Comparison of the three approaches for pangenome assembly.

Approach	Advantages	Limitations
Iterative assembly	Allows construction of a pangenome of a large number of individuals with low (>10×) sequence coverage	Difficulty in genomic placement of novel contigs and identification of copy number variants
Whole genome assembly and comparison	Provides important structural and gene position information	Inability to differentiate between genomic diversity and genome assembly and annotation errors; limited number of individuals due to cost
Graph-based pangenome assembly	Represents the future of pangenome studies	Requires high-quality genome assemblies and substantial computational resources

The programs and packages currently used for pangenome construction and analysis are summarized in Box 1.

reference genome (Bayer et al. 2020). Subsequent mapping of reads from each individual back to the pangenome allows gene PAV calling across the population. Due to the requirement of relatively small quantities of genomic data (studies may use as little as 10× coverage per individual), this approach has been applied for increasingly large population studies, including switchgrass (*Panicum virgatum*) (251 accessions; Lovell et al. 2021), potato (*Solanum* spp.) (296 accessions; Bozan et al. 2023), lettuce (*Lactuca* spp.) (474 accessions; van Workum et al. 2024), and cotton (*Gossypium hirsutum* and *G. barbadense*) (1807 accessions; Li et al. 2021). Despite its broad application, one prerequisite for the iterative approach is the availability of a high-quality reference genome, which may be lacking in many noncrop plant species. In addition, the limitations of the iterative assembly method include the inability to detect copy number variation and the difficulty in placing newly assembled contigs in the reference genome landscape. Given our focus on polyploids here, this method presents the additional challenge of assigning novel contigs to a specific subgenome in an allopolyploid species (formed by hybridization between two species and chromosome doubling). These limitations are being overcome by combining the iterative assembly method with graph pangenomes (see below), using a linearized graph pangenome as a reference for population sequence mapping.

The second approach is the direct comparison of high-quality genome assemblies from multiple individuals in a species. This approach can accurately identify SNPs, small indels, and large SVs, while also providing structural and gene position information (Shi et al. 2023). However, compared to the iterative assembly approach, whole genomes are typically assembled from a relatively small number of individuals due to the high cost of generating high-quality genome assemblies. In addition, differentiating genome assembly and annotation errors from genomic diversity can be challenging (Bayer et al. 2017).

Lastly, with the increasing availability of high-quality genomes and advances in bioinformatics, graph-based approaches are emerging in pangenome studies (e.g. Bayer et al. 2022a, Zanini et al. 2022, Shi et al. 2023, Zhang et al. 2023, Cheng et al. 2025, Li et al. 2025). Pangenome graphs can store all the genetic information of a given species of interest in a graph format, facilitating comparison of genomic variation. Graph-based pangenomes have been used to capture genomic diversity and identify missing heritability, thereby facilitating crop breeding (Edwards and Batley 2022). While the requirements for high-quality genome assemblies and substantial computational resources have previously limited the production of graph pangenomes, this approach, often combined with population-scale iterative mapping, is becoming the standard of the field and represents the future of pangenome studies.

PANGENOMES AND POLYPLOIDY

The main focus of polyploid pangenome research to date has been crop improvement, as most crops are polyploids. For example, Jin et al. (2023) identified 182593 SVs by comparing 11 assembled genomes of polyploid cotton (from *Gossypium hirsutum*, *G. barbadense*, *G. tomentosum*, *G. mustelinum*, and *G. darwinii*). These SV data were then used in GWAS analyses, revealing loci significantly associated with yield and fibre quality traits (Jin et al. 2023). In addition, the SV-based GWAS outperformed the SNP-based GWAS by identifying a greater number of associated loci (Jin et al. 2023). In allotetraploid *Brassica napus*, Song et al. (2020) identified PAVs by comparing eight high-quality reference genomes. PAV-based GWAS identified SVs associated with silique length, seed weight, and flowering time that could not be detected by SNP-based GWAS (Song et al. 2020). As more pangenomes (and graph pangenomes) are constructed, they will become the default reference for genomic studies in polyploid crops, overcoming the limitations of single reference assemblies.

An intriguing finding from polyploid pangenome studies is that polyploid genomes examined to date seem to have a higher proportion of variable genes compared to the diploid parents, although this pattern requires further investigation with larger datasets and more species (Bayer et al. 2020). For example, in *Brassica*, the proportion of variable genes was 38% in the allotetraploid *B. napus*, while the proportions were 21% and 33% in the parental diploid species *B. oleracea* and *B. rapa*, respectively (Bayer et al. 2021). In addition, within *B. napus*, greater genetic diversity was found in the synthetic lines than the natural ones. This finding may be because natural *B. napus* originated from a single polyploidization event, whereas the 20 synthetic lines included in the study were derived from independent crosses between multiple individuals (Bayer et al. 2021).

A graph-based pangenome study in *Cochlearia*, scurvy-grass or spoonwort, a genus of ∼30 species of annual and perennial herbs (Brassicaceae), examined the interactions between WGDs and SVs, revealing that WGDs resulted in a progressive accumulation of deleterious SVs across four ploidal levels (i.e. 2*x*, 4*x*, 6*x*, and 8*x*) (Hämälä et al. 2024). Autotetraploid *Cochlearia* (derived from within-species WGD) carried more SVs than the diploids, and the most prominent difference in SV locations between diploids and tetraploids was the excess of SVs overlapping with exons in the polyploids. Hämälä et al. (2024) argued that the extra allelic copies in the autotetraploid *Cochlearia* masked the recessive mutations, allowing SVs to accumulate. Similarly, the proportion of variable genes increased in tetraploid potato cultivars (88.4%) compared to their diploid wild relatives (74.5%), indicating a more dynamic genome in the polyploids (Sun et al. 2025). However, this pattern is not consistent across all polyploid systems examined. For example, in *Brachypodium*, the D subgenome of the allopolyploid *B. hybridum* did not differ substantially from the diploid parent *B. distachyon* in the proportions of core and variable genes (Gordon et al. 2020). In addition, the *Camelina sativa* (6*x*) pangenome showed lower levels of PAV and SV compared to a closely related diploid, *Arabidopsis thaliana* (Bird et al. 2025). Together, these studies highlight that the effect of polyploidy on genomic variability varies across systems and is lineage-specific.

The mechanism driving gene PAV in polyploids appears to differ from that in diploids. In allopolyploids, nonreciprocal homeologous exchange (HE) between the two subgenomes—where a DNA fragment from one subgenome is replaced by the corresponding homeologous region from the other subgenome—often leads to gene loss (Gaeta and Pires 2010, Wendel et al. 2012). For example, if a gene is present only in one subgenome but not in the other, nonreciprocal HE can replace the gene-containing region with the homeologous region that lacks the gene, resulting in gene

loss (Fig. 1A). Non-reciprocal HE may be a common cause of PAV in polyploid genomes and has been documented in many polyploid systems, including allotetraploid *Brassica napus* (Lloyd *et al.* 2018), synthetic allotetraploid rice (derived from a cross between *Oryza sativa* ssp. *japonica* and *O. sativa* ssp. *indica*) (Li *et al.* 2019), and hexaploid wheat (*Triticum aestivum*) (Heuberger *et al.* 2024). In a pangenome study of allotetraploid *Brassica napus*, analysis of 53 synthetic and nonsynthetic (i.e. cultivated lines) accessions revealed that nonreciprocal HE was the major cause of gene PAV (Hurgobin *et al.* 2018). In addition, compared to the cultivated lines, the synthetics showed greater variation of gene PAV and a higher frequency of HEs (Hurgobin *et al.* 2018). Consistent with these observations, the relatively low level of PAV in the allohexaploid *Camelina sativa* pangenome may result from the paucity of HE (Bird *et al.* 2025).

In contrast, in diploids (and some polyploids), transposable elements (TEs) seem to be related to gene PAV. For example, in diploid *Brassica oleracea*, the density of TEs surrounding variable genes was significantly higher compared to that surrounding core genes (Golicz *et al.* 2016b). Additionally, gene loss modelling indicated that TEs had a greater impact on gene loss in diploid *Brassica* species than in polyploid species (Bayer *et al.* 2021). To identify which genomic features most significantly influence gene loss, Bayer *et al.* (2021) employed a machine learning approach to model gene loss propensity in the pangenomes of *B. oleracea* (2*x*), *B. rapa* (2*x*), and *B. napus* (4*x*). They found that gene loss propensity was associated primarily with TEs in the diploids, whereas nonreciprocal HE was the main driver of gene loss in the polyploid (Bayer *et al.* 2021). Similarly, a pangenome study in the diploid *Brachypodium distachyon* revealed that TE mobilization was an important mechanism in generating gene PAV: the variable genes were found to be located closer to TEs compared to core genes (Gordon *et al.* 2017). Because nonreciprocal HE occurs only in allopolyploids, TEs may also play an important role in generating gene PAV in autopolyploids. For example, in both diploid and autotetraploid *Cochlearia*, TEs were present in ∼60% of SVs, indicating that TE mobilization may be a major cause of SVs (which can lead to gene PAV) regardless of ploidy (Hämälä *et al.* 2024). TEs are also associated with gene PAV in allopolyploid cotton, in which there is scant evidence for nonreciprocal HE (Conover and Wendel 2022). In the pangenomes of both allotetraploid *Gossypium hirsutum* and *G. barbadense*, TE insertion frequency was higher in regions adjacent to variable genes compared to core genes (Li *et al.* 2021).

Pangenome studies provide valuable insights into the evolution of polyploid genomes. According to the gene balance hypothesis (Birchler and Veitia 2007, 2012), genes encoding proteins that act in multiprotein complexes are preferentially retained in duplicate following WGD. As a result, these genes are also more likely to be present across different lineages/accessions of a polyploid species and are identified as core genes. Supporting this hypothesis, in the polyploid *Brassica napus* pangenome (constructed using 79 individuals), a higher proportion of core genes (86%) was involved in protein–protein interaction networks compared to variable genes (72%) (Bayer *et al.* 2021). In addition, the distribution of core and variable genes may differ between subgenomes in polyploids (Fig. 1A). In the palaeohexaploid *Brassica oleracea* pangenome (constructed using nine individuals), the least fractionated subgenome (LF) had the fewest variable genes, whereas the more fractionated subgenome (MF2) contained the most variable genes (Golicz *et al.* 2016b). In an allotetraploid cotton (*Gossypium* spp.) pangenome (containing eight genomes), the density of gene PAVs was higher in the D subgenome than in the A subgenome (Peng *et al.* 2022). Li *et al.* (2021) also found that variable genes from the D subgenome had a faster evolutionary rate than those from the A subgenome. In hexaploid wheat (*Triticum aestivum*), the B subgenome exhibited the highest number of PAVs compared to the A and D subgenomes (Jiao *et al.* 2025). Moreover, pangenome analysis of the diploid parents may signal the fate of genes in a polyploid species. For example, Zhuang *et al.* (2022) constructed a superpangenome in diploid *Glycine* species by integrating 26 genomes. Compared to core genes, variable genes in the diploid parents were more prone to loss in both subgenomes of the naturally occurring tetraploid *Glycine dolichocarpa*: on average, the percentages of gene loss in the polyploid were 44.7% and 65.8% among core and variable genes, respectively (Zhuang *et al.* 2022).

Comparing the pangenomes of polyploids and their progenitors enables the distinction between parental legacy (i.e. the gene PAV patterns in the polyploid are a legacy of gene PAV patterns already present in the progenitor species) and post-WGD genomic changes in shaping the polyploid genome. A pangenome study in *Brachypodium*, which included five allopolyploid *B. hybridum* genomes and 51 genomes from the diploid parent *B. distachyon*, highlighted that the degree of polyploid genome evolution may be overestimated if a single polyploid genome was compared to a single reference genome of the parental species (Gordon *et al.* 2020). This study revealed that the intraspecific variation of the diploid *B. distachyon* explained the patterns of variation found in the corresponding subgenomes in the allopolyploid *B. hybridum*. Therefore, the observed dynamic genome of the polyploid is probably due to parental legacy instead of genomic changes following polyploidy (Gordon *et al.* 2020; Fig. 1B).

THE FUTURE OF PANGENOMICS IN POLYPLOIDS

Future studies should aim to provide more high-quality pangenomes for additional plant species, with more emphasis on naturally occurring species, including polyploids. Although significant progress has been made in plant genome sequencing, there is still a lack of high-quality reference genomes (Kress *et al.* 2022, Sun Y. *et al.* 2022). For example, as of November 2024, the National Center for Biotechnology Information (NCBI) genome dataset (https://www.ncbi.nlm.nih.gov/datasets/genome/) contained 1871 angiosperm reference genomes. However, only 396 (21.2%) of these genomes are chromosome-level assemblies with annotations. Furthermore, due to the increased complexity of polyploid genomes, there are far fewer polyploids with high-quality, chromosome-level reference genomes compared to diploids (Wang Y. *et al.* 2023).

Recent advances in sequencing technologies, including PacBio High-Fidelity (HiFi) long-read sequencing and high-throughput chromatin conformation capture (Hi-C), along with improvements in genome assembly algorithms, have enabled the assembly of high-quality polypoid genomes (Wang Y. *et al.* 2023). Assembling a polyploid reference genome involves multiple steps. First,

in an allopolyploid, assembled contigs must be accurately assigned to specific subgenomes—a process known as subgenome phasing. When progenitor genomes are available, this parental genome information can assist in subgenome phasing (Wang Y. *et al.* 2023). However, if the genomes of the progenitor species are unsequenced, or if the parents are unknown, or extinct, novel computational tools (e.g. SubPhaser and a phylogeny-based approach) have been developed for subgenome assignment (Jia *et al.* 2022, Yan *et al.* 2024). Second, haplotypes need to be resolved within each subgenome in an allopolyploid or across the entire genome in an autopolyploid (Yuan *et al.* 2021b). High similarity among haplotypes in polypoid genomes—often due to identity-by-descent (IBD) issues—poses challenges for accurate haplotype phasing (Kong *et al.* 2023). To address this problem, various strategies have been developed, including reference-based phasing, assembly-based phasing, and gamete binning (Wang Y. *et al.* 2023, Zhang X. *et al.* 2024). In reference-based phasing, sequencing reads are aligned to a reference genome to identify variant sites, which are then used to distinguish different haplotypes. This method was used to generate a haplotype-resolved genome of hexaploid sweet potato (*Ipomoea batatas*) (Yang *et al.* 2017). Assembly-based phasing utilizes haplotype-aware assemblers (e.g. hifiasm) and haplotype-aware scaffolders (e.g. YaHS and ALLHiC) to construct haplotype-resolved polyploid genomes, as demonstrated in tetraploid *Actinidia arguta* (Zhang F. *et al.* 2024) and *Urochloa decumbens* (4*x*) (Ryan *et al.* 2025). Gamete binning utilizes single-cell sequencing of gametic genomes to group contigs by haplotype, followed by scaffolding. This approach was used in assembling the autotetraploid potato (*Solanum tuberosum*) genome (Sun H. *et al.* 2022).

Following genome assembly, the annotation consistency and quality also need to be considered in pangenome studies. Weisman *et al.* (2022) found that mixed annotation processes can lead to misidentification of lineage-specific genes. Compared to using a uniform annotation approach across related species, the use of different annotation sources resulted in up to a 15-fold change in the number of lineage-specific genes (Weisman *et al.* 2022).

High-quality pangenomes, consisting of subgenome-phased, haplotype-resolved, and uniformly annotated polyploid genomes, will enable accurate and comprehensive identification of structural variation both within a single polyploid genome (i.e. variation among haplotypes and subgenomes) and across individuals and populations. A polyploid species, with these multiple layers of genetic variation, may use this diversity as a source for adaptation, stress response, and the evolution of novel phenotypes.

As reviewed here, most current pangenome studies focus on crops and their wild relatives aimed at supporting crop improvement. However, pangenomes in natural polyploid populations are also promising resources for ecological and evolutionary studies (Schreiber *et al.* 2024). One of the few pangenomic studies on naturally occurring polyploids is that of *Cochlearia* (Brassicaceae) noted above. Hämälä *et al.* (2024) found that SVs play diverse and contrasting roles in the evolution of autotetraploid *Cochlearia*. Although polyploids accumulated more deleterious SVs than diploids, the increased SV loads in polyploids also led to a greater diversity of adaptive alleles, potentially contributing to the local adaptation of the polyploids (Hämälä *et al.* 2024). In other natural polyploid systems, the acquisition or loss of genes during and following WGD may also partially explain the adaptability and evolutionary success of polyploids, shedding light on how certain polyploid lineages adapt to new environmental conditions. Analysis of the features of variable genes highlights the role of local transposons in gene loss and the impact of nonreciprocal HEs on variable gene content in polyploids (Bayer *et al.* 2021). This genomic diversity, when under selection, offers significant potential for adaptation and may be one of the major reasons why polyploidy is so common among crop species, as well as plants in natural populations. Future research comparing the genomes of different naturally occurring polyploid individuals could reveal diverse evolutionary trajectories of polyploid genomes and their associations with local environments, providing valuable insights for both basic and applied plant science.

In addition, there are a few recently formed (<200 years old) natural polyploids with clear parentage, including species in *Spartina*, *Senecio*, *Cardamine*, *Tragopogon*, and *Mimulus* (Soltis and Soltis 2009, Wendel *et al.* 2018, Edger *et al.* 2025). These systems provide unique opportunities to study the immediate consequences of WGD in nature. However, no pangenomic studies have been conducted on any of these species to date. Pangenome-level comparisons between polyploids and their diploid progenitors will provide novel insight into polyploid formation and genome evolution. The parentage and independent origins of the polyploids could be systematically examined in greater detail through pangenomic analysis of both diploids and polyploids. For example, if an SV is unique to a specific diploid lineage, the polyploid individual containing the same SV is very likely a descendant of that diploid. Associating a polyploid with a diploid lineage through these inherited SVs could be helpful in identifying independent origins of polyploids. In addition, the pangenomes of independently formed polyploids may reveal both unique and shared patterns of genome evolution immediately following polyploidy. This knowledge will shed light on how polyploids adapt to local environments and could facilitate the breeding of crops through polyploidization.

Whether polyploid pangenomes generally contain a higher proportion of variable genes than those of the diploid parents remains unresolved. How ploidal level affects genome variability needs further investigation. The answer to this question may provide valuable insights for crop improvement; the dynamic nature of polyploid genomes may facilitate the generation of beneficial mutants and promote plant breeding efforts. Furthermore, such data would be very useful in improving our understanding of the success of polyploids in nature. On the one hand, the genetic bottleneck associated with polyploid formation reduces the amount of genetic variation (including PAV and SV) in the newly formed polyploid compared to that present in the diploid parents (Udall and Wendel 2006, Soltis *et al.* 2014, Levin 2019). On the other hand, multiple origins of polyploids are prevalent among plant systems, and these will incorporate more genetic diversity in the ancestral polyploids (Soltis *et al.* 2004). In addition, because of the presence of the extra genome, and therefore genetic load, relaxed purifying selection on deleterious recessive mutations in polyploids may contribute to increased variability in the pangenome, and the dynamic nature of genome evolution in polyploids,

including gene loss, HE, sub-/neofunctionalization, could incorporate additional genetic changes to the genomes.

How allopolyploid pangenomes compare to autopolyploid pangenomes remains unstudied. Given that nonreciprocal HE, a major source of gene PAV, occurs only in allopolyploids (because autopolyploids, by definition, lack subgenomes), one may ask: does the allopolyploid pangenome typically contain a higher proportion of variable genes than that of an autopolyploid? Hoopes *et al.* (2022) constructed a pangenome in autotetraploid potato (*Solanum tuberosum*) using six genomes and found that 21.2% of the genes were variable. In comparison, pangenome studies of allotetraploids with similar sample size reported higher proportions of variable genes: 68.5% and 43.6% in *Brassica napus* (nine genomes) and *Gossypium hirsutum* (15 individuals), respectively (Song *et al.* 2020, Li *et al.* 2024). However, more polyploid systems are needed for a more comprehensive survey, as are more pangenomic studies of related autopolyploids and allopolyploids (e.g. from the same genus or family).

Methodologically, pangenome-level comparisons between diploids and polyploids, and between allopolyploids and autopolyploids, may be constrained by various factors, including sample size and the genetic relatedness of the accessions or cultivars included in the study (Montenegro *et al.* 2017). The larger the sample size, the smaller the proportion of core genes observed (Shi *et al.* 2023). A solution to eliminate the bias due to sample size is to randomly select a subset of individuals from the larger population, matching the number of individuals in the smaller population (Zhuang *et al.* 2022). Future systematic pangenome-level comparisons between diploids and polyploids, as well as between allopolyploids and autopolyploids, utilizing well-matched sample populations, will be important to accurately assess how ploidal level and polyploid type influence pangenome variability.

Pangenomes can help illuminate how SVs influence polyploid formation. In newly formed autopolyploids, multivalent pairing can lead to abnormal segregation and produce nonviable, aneuploid gametes, and early selection against meiotic instability is thought to be important for the long-term success in autopolyploids that mainly reproduce sexually (Bomblies 2023). The presence of SVs between two divergent individuals of a species that interbreed (e.g. following long-distance dispersal), followed by WGD in their progeny, could lead to pairing bias between structurally similar chromosomes and increased meiotic stability. Although this effect is likely to be transient and eventually superseded by other regulatory mechanisms in autopolyploids (Bomblies *et al.* 2016, Bomblies 2023), any selective advantage during the perilous first generations of a polyploid lineage could be relevant to their survival. Future studies could use diploid pangenomes from an autopolyploid progenitor to identify lines with significant structural differences, produce synthetic polyploids from crosses within and between those lines, and compare their meiotic characteristics and fertility (e.g. Parra-Nunez *et al.* 2019).

For allopolyploids, SVs could influence the compatibility of the genomes of two species during hybridization. For example, interactions between variants that result in inviability or complete infertility at the homoploid level could prevent the formation of an intermediate diploid F_1 hybrid, a common first step in allopolyploid formation (Tayalé and Parisod 2013). Some repeatedly formed allopolyploids show a bias toward certain combinations of genotypes in the diploid progenitor species (Soltis *et al.* 2023), and investigation of allopolyploid and diploid pangenomes could address whether SVs influence this bias.

The genomes of newly formed polyploids are often destabilized following WGD—aneuploidy, TE activation, and homeologous recombination are frequently observed in resynthesized and very young polyploid lineages (e.g. Ramsey and Schemske 2002, Tayalé and Parisod 2013). These phenomena could lead to a rapid increase in the number of SVs within the pangenome of a newly formed polyploid lineage over several generations. As noted above, the application of pangenome analyses to polyploids and their parents can help discern which SVs in the polyploid are inherited from the diploid and which arose after WGD. Another question is how much of the polyploid's pangenome diversity accumulated early after WGD, and whether certain types or patterns of SVs are preferentially retained. Research along these lines would build on the study of homeologue loss and biased fractionation in allopolyploids, where the types and parental origins of genes may significantly affect their fate following WGD (Wendel *et al.* 2018). Methodologically, it will be important to address whether current methods can differentiate SVs that arose within a lineage from those inherited from past introgression from independent polyploid lineages (if they overlap) or even diploid progenitor species via unreduced gametes. Thorough sampling will certainly be needed for any pangenome investigation of these questions at a large scale.

In summary, pangenome studies in polyploids hold enormous potential in agriculture as well as ecological and evolutionary studies. Generating high-quality polyploid pangenomes will enable a more comprehensive understanding of genetic variation that drives the evolution of novel traits and adaptations. Pangenomic studies will also provide novel insights into the formation and evolutionary trajectories of polyploids. Additionally, further investigation is needed to explore how ploidal level and polyploid type (e.g. autopolyploid and the various intermediates to allopolyploid) impact pangenome variability. We hope this article inspires the growth of pangenomic research in polyploid systems, particularly within natural populations.

AUTHOR CONTRIBUTIONS

Shengchen Shan: Conceptualization (Equal); writing—original draft (Equal); writing—review and editing (Equal); Jonathan P. Spoelhof: Conceptualization (Equal); writing—original draft (Supporting); writing—review and editing (Supporting); Paul D. Blischak: Conceptualization (Equal); writing—original draft (Supporting); writing—review and editing (Supporting); Jacqueline Batley: Conceptualization (Equal); writing—original draft (Supporting); writing—review and editing (Supporting); Pamela S. Soltis: Conceptualization (Equal); writing—original draft (Supporting); writing—review and editing (Supporting); Douglas E. Soltis: Conceptualization (Equal); writing—original draft (Supporting); writing—review and editing (Supporting); and David Edwards: Conceptualization (Equal); writing—original draft (Supporting); writing—review and editing (Supporting).

SUPPLEMENTARY DATA

Supplementary data are available at *Evolutionary Journal of the Linnean Society* online.

CONFLICT OF INTEREST

None declared.

FUNDING

This work was supported by US National Science Foundation grants IOS-1923234 and DEB-2043478 to D.E.S. and P.S.S. and DBI-2320251 to P.S.S. and D.E.S..

REFERENCES

Bayer PE, Golicz AA, Scheben A *et al.* Plant pan-genomes are the new reference. *Nature Plants* 2020;**6**:914–20.

Bayer PE, Hurgobin B, Golicz AA *et al.* Assembly and comparison of two closely related *Brassica napus* genomes. *Plant Biotechnology Journal* 2017;**15**:1602–10.

Bayer PE, Petereit J, Durant É *et al.* Wheat Panache: a pangenome graph database representing presence–absence variation across sixteen bread wheat genomes. *The Plant Genome* 2022a;**15**:e20221.

Bayer PE, Scheben A, Golicz AA *et al.* Modelling of gene loss propensity in the pangenomes of three *Brassica* species suggests different mechanisms between polyploids and diploids. *Plant Biotechnology Journal* 2021;**19**:2488–500.

Bayer PE, Valliyodan B, Hu H *et al.* Sequencing the USDA core soybean collection reveals gene loss during domestication and breeding. *The Plant Genome* 2022b;**15**:e20109.

Birchler JA, Veitia RA. The gene balance hypothesis: from classical genetics to modern genomics. *The Plant Cell* 2007;**19**:395–402.

Birchler JA, Veitia RA. Gene balance hypothesis: connecting issues of dosage sensitivity across biological disciplines. *Proceedings of the National Academy of Sciences of the United States of America* 2012;**109**:14746–53.

Bird KA, Brock JR, Grabowski PP *et al.* Allopolyploidy expanded gene content but not pangenomic variation in the hexaploid oilseed *Camelina sativa*. *Genetics* 2025;**229**:iyae183.

Bomblies K. Learning to tango with four (or more): the molecular basis of adaptation to polyploid meiosis. *Plant Reproduction* 2023;**36**:107–24.

Bomblies K, Jones G, Franklin C *et al.* The challenge of evolving stable polyploidy: could an increase in "crossover interference distance" play a central role?*Chromosoma* 2016;**125**:287–300.

Bozan I, Achakkagari SR, Anglin NL *et al.* Pangenome analyses reveal impact of transposable elements and ploidy on the evolution of potato species. *Proceedings of the National Academy of Sciences of the United States of America* 2023;**120**:e2211117120.

Cheng L, Wang N, Bao Z *et al.* Leveraging a phased pangenome for haplotype design of hybrid potato. *Nature* 2025;**640**:408–17.

Conover JL, Wendel JF. Deleterious mutations accumulate faster in allopolyploid than diploid cotton (*Gossypium*) and unequally between subgenomes. *Molecular Biology and Evolution* 2022;**39**: msac024.

Cortinovis G, Vincenzi L, Anderson R *et al.* Adaptive gene loss in the common bean pan-genome during range expansion and domestication. *Nature Communications* 2024;**15**:6698.

Edger PP, Soltis DE, Yoshioka S *et al.* Natural neopolyploids: a stimulus for novel research. *The New Phytologist* 2025;**246**:78–93.

Edwards D, Batley J. Graph pangenomes find missing heritability. *Nature Genetics* 2022;**54**:919–20.

Gaeta RT, Pires JC. Homoeologous recombination in allopolyploids: the polyploid ratchet. *The New Phytologist* 2010;**186**:18–28.

Golicz AA, Batley J, Edwards D. Towards plant pangenomics. *Plant Biotechnology Journal* 2016a;**14**:1099–105.

Golicz AA, Bayer PE, Barker GC *et al.* The pangenome of an agronomically important crop plant *Brassica oleracea*. *Nature Communications* 2016b;**7**:13390.

Golicz AA, Bayer PE, Bhalla PL *et al.* Pangenomics comes of age: from bacteria to plant and animal applications. *Trends in Genetics: TIG* 2020;**36**:132–45.

Gordon SP, Contreras-Moreira B, Levy JJ *et al.* Gradual polyploid genome evolution revealed by pan-genomic analysis of *Brachypodium hybridum* and its diploid progenitors. *Nature Communications* 2020;**11**:3670.

Gordon SP, Contreras-Moreira B, Woods DP *et al.* Extensive gene content variation in the *Brachypodium distachyon* pan-genome correlates with population structure. *Nature Communications* 2017;**8**:2184.

Guo L, Wang X, Ayhan DH *et al.* Super pangenome of *Vitis* empowers identification of downy mildew resistance genes for grapevine improvement. *Nature Genetics* 2025;**57**:741–53.

Hämälä T, Moore C, Cowan L *et al.* Impact of whole-genome duplications on structural variant evolution in *Cochlearia*. *Nature Communications* 2024;**15**:5377.

He F, Chen S, Zhang Y *et al.* Pan-genomic analysis highlights genes associated with agronomic traits and enhances genomics-assisted breeding in alfalfa. *Nature Genetics* 2025;**57**:1262–73. https://doi.org/10.1038/s41588-025-02164-8.

Heuberger M, Bernasconi Z, Said M *et al.* Analysis of a global wheat panel reveals a highly diverse introgression landscape and provides evidence for inter-homoeologue chromosomal recombination. *TAG. Theoretical and Applied Genetics. Theoretische und Angewandte Genetik* 2024;**137**:236.

Hoopes G, Meng X, Hamilton JP *et al.* Phased, chromosome-scale genome assemblies of tetraploid potato reveal a complex genome, transcriptome, and predicted proteome landscape underpinning genetic diversity. *Molecular Plant* 2022;**15**:520–36.

Hu H, Scheben A, Verpaalen B *et al. Amborella* gene presence/absence variation is associated with abiotic stress responses that may contribute to environmental adaptation. *The New Phytologist* 2022;**233**:1548–55.

Hurgobin B, Golicz AA, Bayer PE *et al.* Homoeologous exchange is a major cause of gene presence/absence variation in the amphidiploid *Brassica napus*. *Plant Biotechnology Journal* 2018;**16**:1265–74.

Jaggi KE, Krak K, Štorchová H *et al.* A pangenome reveals LTR repeat dynamics as a major driver of genome evolution in *Chenopodium*. *The Plant Genome* 2025;**18**:e70010.

Jia KH, Wang ZX, Wang L *et al.* SubPhaser: a robust allopolyploid subgenome phasing method based on subgenome-specific *k*-mers. *The New Phytologist* 2022;**235**:801–9.

Jiao C, Xie X, Hao C *et al.* Pan-genome bridges wheat structural variations with habitat and breeding. *Nature* 2025;**637**:384–93.

Jiao WB, Schneeberger K. Chromosome-level assemblies of multiple *Arabidopsis* genomes reveal hotspots of rearrangements with altered evolutionary dynamics. *Nature Communications* 2020;**11**:989.

Jin S, Han Z, Hu Y *et al.* Structural variation (SV)-based pan-genome and GWAS reveal the impacts of SVs on the speciation and diversification of allotetraploid cottons. *Molecular Plant* 2023;**16**:678–93.

Kang M, Wu H, Liu H *et al.* The pan-genome and local adaptation of *Arabidopsis thaliana*. *Nature Communications* 2023;**14**:6259.

Khan AW, Garg V, Sun S *et al. Cicer* super-pangenome provides insights into species evolution and agronomic trait loci for crop improvement in chickpea. *Nature Genetics* 2024;**56**:1225–34.

Kong W, Wang Y, Zhang S *et al.* Recent advances in assembly of complex plant genomes. *Genomics, Proteomics & Bioinformatics* 2023;**21**:427–39.

Kress WJ, Soltis DE, Kersey PJ *et al.* Green plant genomes: What we know in an era of rapidly expanding opportunities. *Proceedings of the National Academy of Sciences* 2022;**119**:e2115640118.

Lei L, Goltsman E, Goodstein D *et al.* Plant pan-genomics comes of age. *Annual Review of Plant Biology* 2021;**72**:411–35.

Levin DA. Why polyploid exceptionalism is not accompanied by reduced extinction rates. *Plant Systematics and Evolution* 2019;**305**:1–11.

Li J, Liu Z, You C *et al.* Convergence and divergence of diploid and tetraploid cotton genomes. *Nature Genetics* 2024;**56**:2562–73.

Li J, Yuan D, Wang P *et al.* Cotton pan-genome retrieves the lost sequences and genes during domestication and selection. *Genome Biology* 2021;**22**:119.
Li N, Xu C, Zhang A *et al.* DNA methylation repatterning accompanying hybridization, whole genome doubling and homoeolog exchange in nascent segmental rice allotetraploids. *The New Phytologist* 2019;**223**:979–92.
Li W, Chu C, Zhang T *et al.* Pan-genome analysis reveals the evolution and diversity of *Malus*. *Nature Genetics* 2025;**57**:1274–86. https://doi.org/10.1038/s41588-025-02166-6.
Li YH, Zhou G, Ma J *et al. De novo* assembly of soybean wild relatives for pan-genome analysis of diversity and agronomic traits. *Nature Biotechnology* 2014;**32**:1045–52.
Lian Q, Huettel B, Walkemeier B *et al.* A pan-genome of 69 *Arabidopsis thaliana* accessions reveals a conserved genome structure throughout the global species range. *Nature Genetics* 2024;**56**:982–91.
Lloyd A, Blary A, Charif D *et al.* Homoeologous exchanges cause extensive dosage-dependent gene expression changes in an allopolyploid crop. *The New Phytologist* 2018;**217**:367–77.
Lovell JT, MacQueen AH, Mamidi S *et al.* Genomic mechanisms of climate adaptation in polyploid bioenergy switchgrass. *Nature* 2021;**590**:438–44.
Montenegro JD, Golicz AA, Bayer PE *et al.* The pangenome of hexaploid bread wheat. *The Plant Journal: For Cell and Molecular Biology* 2017;**90**:1007–13.
Parra-Nunez P, Pradillo M, Santos JL. Competition for chiasma formation between identical and homologous (but not identical) chromosomes in synthetic autotetraploids of *Arabidopsis thaliana*. *Frontiers in Plant Science* 2019;**9**:1924.
Peng R, Xu Y, Tian S *et al.* Evolutionary divergence of duplicated genomes in newly described allotetraploid cottons. *Proceedings of the National Academy of Sciences of the United States of America* 2022;**119**:e2208496119.
Ramsey J, Schemske DW. Neopolyploidy in flowering plants. *Annual Review of Ecology and Systematics* 2002;**33**:589–639.
Raza A, Bohra A, Garg V *et al.* Back to wild relatives for future breeding through super-pangenome. *Molecular Plant* 2023;**16**:1363–5.
Rijzaani H, Bayer PE, Rouard M *et al.* The pangenome of banana highlights differences between genera and genomes. *The Plant Genome* 2022;**15**:e20100.
Ryan C, Fraser F, Irish N *et al.* A haplotype-resolved chromosome-level genome assembly of *Urochloa decumbens* cv. Basilisk resolves its allopolyploid ancestry and composition. *G3: Genes, Genomes, Genetics* 2025;**15**: jkaf005.
Schreiber M, Jayakodi M, Stein N *et al.* Plant pangenomes for crop improvement, biodiversity and evolution. *Nature Reviews. Genetics* 2024;**25**:563–77.
Shi J, Tian Z, Lai J *et al.* Plant pan-genomics and its applications. *Molecular Plant* 2023;**16**:168–86.
Soltis DE, Soltis PS, Tate JA. Advances in the study of polyploidy since plant speciation. *New Phytologist* 2004;**161**:173–91.
Soltis DE, Tate JA, Soltis PS *et al.* Parental genomic compatibility model: only certain diploid genotype combinations form allopolyploids. *The Nucleus* 2023;**66**:371–8.
Soltis DE, Visger CJ, Soltis PS. The polyploidy revolution then … and now: Stebbins revisited. *American Journal of Botany* 2014;**101**:1057–78.
Soltis PS, Soltis DE. The role of hybridization in plant speciation. *Annual Review of Plant Biology* 2009;**60**:561–88.
Song B, Ning W, Wei D *et al.* Plant genome resequencing and population genomics: current status and future prospects. *Molecular Plant* 2023;**16**:1252–68.
Song JM, Guan Z, Hu J *et al.* Eight high-quality genomes reveal pan-genome architecture and ecotype differentiation of *Brassica napus*. *Nature Plants* 2020;**6**:34–45.
Sun H, Jiao WB, Krause K *et al.* Chromosome-scale and haplotype-resolved genome assembly of a tetraploid potato cultivar. *Nature Genetics* 2022;**54**:342–8.
Sun H, Tusso S, Dent CI *et al.* The phased pan-genome of tetraploid European potato. *Nature* 2025;**642**:389–97. https://doi.org/10.1038/s41586-025-08843-0.
Sun Y, Shang L, Zhu QH *et al.* Twenty years of plant genome sequencing: achievements and challenges. *Trends in Plant Science* 2022;**27**:391–401.
Tayalé A, Parisod C. Natural pathways to polyploidy in plants and consequences for genome reorganization. *Cytogenetic and Genome Research* 2013;**140**:79–96.
Tettelin H, Masignani V, Cieslewicz MJ *et al.* Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: Implications for the microbial "pan-genome." *Proceedings of the National Academy of Sciences of the United States of America* 2005;**102**:13950–5.
Udall JA, Wendel JF. Polyploidy and crop improvement. *Crop Science* 2006;**46**:S3–S14.
van Workum DJM, Mehrem SL, Snoek BL *et al. Lactuca* super-pangenome reduces bias towards reference genes in lettuce research. *BMC Plant Biology* 2024;**24**:1019.
Wang N, Li Y, Shen C *et al.* High-resolution sequencing of nine elite upland cotton cultivars uncovers genic variations and breeding improvement targets. *The Plant Journal: For Cell and Molecular Biology* 2023;**113**:145–59.
Wang Y, Yu J, Jiang M *et al.* Sequencing and assembly of polyploid genomes. In: Van de Peer Y (ed.), *Polyploidy: Methods and Protocols*. New York: Humana Press, 2023, 429–58.
Wendel JF, Flagel LE, Adams KL, Jeans, genes, and genomes: cotton as a model for studying polyploidy. In: Soltis PS, Soltis DE (eds.), *Polyploidy and Genome Evolution*. Berlin: Springer, 2012, 181–207.
Wendel JF, Lisch D, Hu G *et al.* The long and short of doubling down: Polyploidy, epigenetics, and the temporal dynamics of genome fractionation. *Current Opinion in Genetics & Development* 2018;**49**:1–7.
Weisman CM, Murray AW, Eddy SR. Mixing genome annotation methods in a comparative analysis inflates the apparent number of lineage-specific genes. *Current Biology: CB* 2022;**32**:2632–9.e2.
Yan X, Shi G, Sun M *et al.* Genome evolution of the ancient hexaploid *Platanus* × *acerifolia* (London planetree). *Proceedings of the National Academy of Sciences of the United States of America* 2024;**121**:e2319679121.
Yang J, Moeinzadeh MH, Kuhl H *et al.* Haplotype-resolved sweet potato genome traces back its hexaploidization history. *Nature Plants* 2017;**3**:696–703.
Yuan Y, Bayer PE, Batley J *et al.* Current status of structural variation studies in plants. *Plant Biotechnology Journal* 2021a;**19**:2153–63.
Yuan Y, Scheben A, Edwards D *et al.* Toward haplotype studies in polyploid plants to assist breeding. *Molecular Plant* 2021b;**14**:1969–72.
Zanini SF, Bayer PE, Wells R *et al.* Pangenomics in crop improvement—from coding structural variations to finding regulatory variants with pangenome graphs. *The Plant Genome* 2022;**15**:e20177.
Zhang F, Wang Y, Lin Y *et al.* Haplotype-resolved genome assembly provides insights into evolutionary history of the *Actinidia arguta* tetraploid. *Molecular Horticulture* 2024;**4**:4.
Zhang X, Chen Y, Wang L *et al.* Pangenome of water caltrop reveals structural variations and asymmetric subgenome divergence after allopolyploidization. *Horticulture Research* 2023;**10**:uhad203.
Zhang X, Li D, Pan W. Haplotype-resolved assembly of auto-polyploid genomes via combining Hi-C and gametic data. *Scientific Reports* 2024;**14**:7892.
Zhou Y, Zhang Z, Bao Z *et al.* Graph pangenome captures missing heritability and empowers tomato breeding. *Nature* 2022;**606**:527–34.
Zhuang Y, Wang X, Li X *et al.* Phylogenomics of the genus *Glycine* sheds light on polyploid evolution and life-strategy transition. *Nature Plants* 2022;**8**:233–44.