GENOME ANALYSES AND BREEDING OF POLYPLOID CROPS

Lin Cheng[1,2,3,9], Zhigui Bao[4,9], Qianqian Kong[1], Ludivine Lassois[3], Nils Stein[5,6], Sanwen Huang[2,7] & Qian Zhou[1,8]

[1]School of Agriculture and Biotechnology, Sun Yat-Sen University, Shenzhen, China. [2]National Key Laboratory of Tropical Crop Breeding, Shenzhen Branch, Guangdong Laboratory of Lingnan Modern Agriculture, Genome Analysis Laboratory of the Ministry of Agriculture and Rural Affairs, Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen, China. [3]Plant Genetics and Rhizosphere Processes Laboratory, TERRA Teaching and Research Center, University of Liège, Gembloux Agro-Bio Tech, Gembloux, Belgium. [4]Department of Molecular Biology, Max Planck Institute for Biology Tübingen, Tübingen, Germany. [5]Department of Genebank, Leibniz Institute of Plant Genetics and Crop Plant Research (IPK) Gatersleben, Seeland, Germany. [6]Crop Plant Genetics, Institute of Agricultural and Nutritional Sciences, Martin-Luther-University of Halle-Wittenberg, Halle (Saale), Germany. [7]National Key Laboratory of Tropical Crop Breeding, Chinese Academy of Tropical Agricultural Sciences, Haikou, China. [8]Department of Network Intelligence, Peng Cheng Laboratory, Shenzhen, China. [9]These authors contributed equally: Lin Cheng, Zhigui Bao.

ABSTRACT

Polyploidization is a common and important evolutionary process in the plant kingdom. Compared with diploid plant species, the intricate genome architecture of polyploid plant species presents substantial challenges in applying multi-omics approaches for crop breeding improvement. In this Review, we summarize the current techniques for analysing polyploid genomes, including constructing reference genomes and pan-genomes, and detecting variants. We also assess findings related to polyploid genome architecture, population genetics and breeding programmes, highlighting advanced techniques in the breeding of polyploid crops. Finally, we explore the challenges and demands posed by polyploid genome complexity during analysis with available biotechnological tools. This Review emphasizes the importance of a comprehensive understanding of polyploid genomic features for the further genetic improvement of polyploid crops.

Polyploid crops are diverse and widespread, tightly woven into the fabric of agriculture and human civilization. Cytological, fossil and genomic evidence indicates that more than 47% of flowering plant species can be linked to polyploid events during the diversification ofthe angiosperm crown group[1–3]. Of the top 20 crops worldwide based on production tonnage, 6 are polyploid and together account for 48% ofglobal production (Fig. 1 and Supplementary Table 1; https://www.fao.org/faostat/), with hexaploid sugarcane (*Saccharum officinarum*) having the highest global tonnage harvest, representing 80% ofthe world's sugar production[4,5]. Hexaploid bread wheat (*Triticum aestivum* L.) is a key staple grain crop rich in protein, carbohydrates and minerals, accounting for nearly 30% of human calorie intake[6]. Tetraploid potatoes (*Solanum tuberosum* L.) are a staple crop for approximately 1.3 billion people worldwide[7]. Octoploid strawberries (*Fragaria* sp.), allopolyploid *Brassica napus* and cotton (*Gossypium* sp.), among other diverse polyploid crops, provide essential nutrients and resources and contribute to food security and economic development.

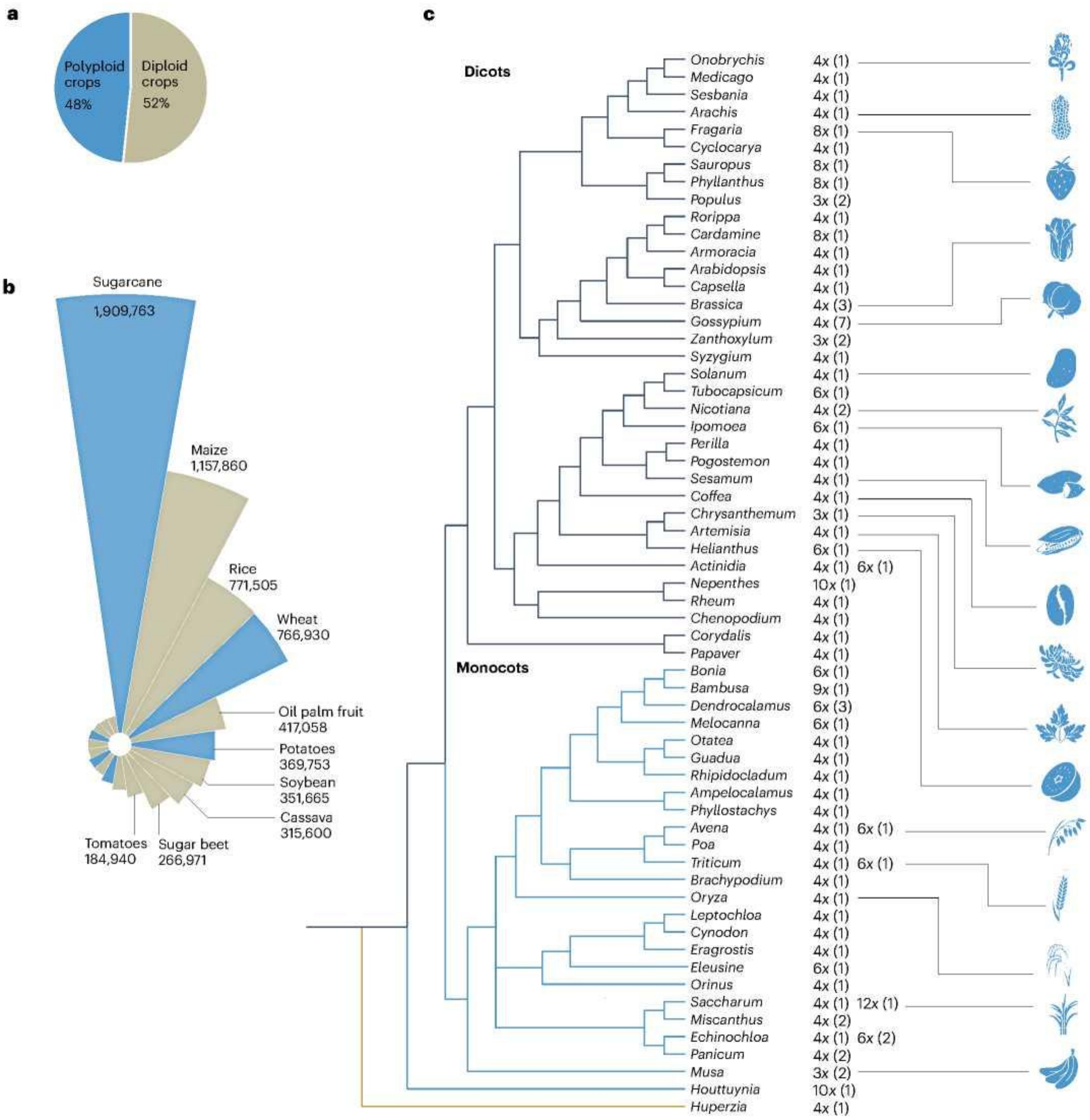
Polyploid crops often show increased vigour and, in some cases, may outperform their diploid counterparts in various traits, perhaps because the multiple sets of chromosomes can buffer mutations and provide added genetic diversity[8]. However, breeding polyploid crops is much more complex than breeding diploid crops because the multiple sets of chromosomes complicate genetic interactions and inheritance patterns[9,10]. Considerable effort is needed to account for the wider range ofgenetic variation in polyploid crop species, which makes selection more challenging. In addition, although the more complex polyploid genome architectures may allow the plants to tolerate more deleterious mutations[11,12], they can also mask recessive traits, thereby making it difficult to stabilize certain traits over generations.

Genome sequencing and analysis are fundamental for understanding and using the characteristics of polyploid crops in functional research and breeding programmes. In general, polyploids can be classified as autopolyploids, which arise from chromosome duplication within a single species, or allopolyploids, which are formed through hybridization between different species[13,14]. In polyploid genomes, distinct sets of DNA sequences are classified on the basis of their evolutionary origins: the term 'haplotypes' commonly refers to allelic sequences inherited from the same ancestral species, whereas 'subgenomes' represent complete chromosome sets derived from different progenitor species. For example, the octoploid cultivated strawberry (*Fragaria* × *ananassa*, $2n = 8x = 56$) has a subgenomic constitution denoted as $A_1A_2B_1B_2C_1C_2D_1D_2$, where the letters A, B, C and D represent 4 divergent subgenomes originating from distinct diploid ancestors, and the numerical suffixes (1 and 2) indicate haplotypes within each subgenome (for example, A_1 and A_2 are 2 haplotypes of the A subgenome)[15,16].

Recent advances in sequencing technologies have helped resolve the detailed architecture of several polyploid genomes, revealing gene loss and acquisition across subgenomes. However, the genomes of most polyploid plant species are challenging to assemble given the limitations of existing genomic techniques, which were developed for assembling diploid genomes. Consequently, the utilization of genomics resources and functional genes in polyploid crops currently lags behind that in diploid crops, a situation that has constrained breeding and improvement of polyploid species. To address this bottleneck, researchers are deploying advanced biotechnologies to rapidly domesticate wild plant species with polyploid genomes or create new polyploids. These technologies are expected to enable genomics-driven rapid breeding, tapping into the agricultural potential of polyploid species to ensure future food security.

In this Review, we provide an overview of the genomic landscape of polyploid crops, focusing on diverse genome sequences, variations and pan-genome research. We also summarize advances in population genetics and linkage mapping in polyploid crops. Finally, we highlight modern technologies such as genome editing, genome selection and synthetic breeding that are being used to accelerate breeding. We anticipate that such programmes will pave the way for the rapid development and valorization of polyploid crops.

Fig. 1 | Global yield and phylogeny of sequenced polyploid crops. a, Yield production of polyploid and diploid crops. These statistics are based on the sum of the average production of the top 20 crops in 2018–2022, as released by the United Nations Food and Agriculture Organization (https://www.fao. org/faostat/). **b**, Yield production (in tons) ranking of the top 20 crops, listed from highest to lowest (Supplementary Table 1). Note: only the top 10 crops are labelled. **c**, Phylogenetic tree of 61 genera containing polyploid species, developed using TimeTree (https://timetree.org/)[180]. The labels (number + *x*) in the tree indicate the ploidy level of representative polyploid species in each genus that we collected; the number in brackets indicates the number of species with available sequenced genomes collected from Web of Science (https://www. webofscience.com/wos) and PubPlant (https://www.plabipd.de/pubplant_main. html, bioRxiv excluded) around May 2024 (Supplementary Table 2). In the phylogenetic tree, branches coloured brown correspond to the outgroup genus, those coloured blue correspond to monocots and those coloured black correspond to dicots.



Published in : Nature Plants
DOI: 10.1038/s41477-025-02088-5
Status : Postprint (Author's version)

Genome organization of polyploid crops

Assembly and annotation of polyploid genomes

Polyploid plant species typically have multiple chromosome sets with highly similar sequences, which pose additional challenges for their assembly relative to that of genomes from homozygous inbred lines or diploid species. Early efforts to assemble polyploid genomes primary focused on major crops, such as rapeseed (*B. napus*)[17], cotton (*Gossypium hirsutum*)[18] and bread wheat[6]. These assemblies relied on sequencing bacterial artificial chromosomes and multiple types of short read, combined with linkage maps, often requiring consortium-level efforts. With the advance of long-read sequencing, the landscape of polyploid genome assembly has drastically improved (Table 1). The high density of genetic variants within subgenomes, arising from the hybridization of distinct ancestral species, facilitates differentiation using long-read platforms, even those with relatively high error rates, such as PacBio continuous long reads and Oxford Nanopore Technologies R9 simplex reads. High-accuracy long reads and high-performance assembly algorithms[19,20] have enhanced the assembly quality of complex polyploid genomes, enabling the distinct identification of subgenomes and accurate assembly of haplotypes in species such as potato[21,22], strawberry[23] and Ma bamboo (*Dendrocalamus latiflorus*)[24].

The earliest and most widely used method to distinguish between subgenomes relies on information from the diploid ancestors of polyploid species[1,25–27], such as rapeseed[17], peanut (*Arachis hypogaea*)[27] and bread wheat[25]. The evolutionary signals that persist after allopolyploidy could be recorded by the presence of subgenome-specific transposons[28], or tandem repeats[29], along with a more general method using specific *k*-mers, which provide alternative methods for identifying subgenomes[30,31]. However, frequent ancestral recombination between subgenomes[32,33] in some allopolyploids, such as *Chrysanthemum morifolium* and Chinese mugwort (*Artemisia argyi*), makes subgenome differentiation more challenging[34].

The assembly of autopolyploid genomes lags behind that of allopolyploids. The first attempt at generating haplotype-resolved (allele-aware) assemblies for an autopolyploid plant species was for the sweet potato (*Ipomoea batatas*) genome, using local phasing of read-based variant calling[35]. Subsequently, the ALLHiC algo- rithm[36] was developed to assemble autopolyploid genomes based on high-throughput chromosome conformation capture (Hi-C) data, using diploid reference genome sequences to remove redundant signals from different haplotypes. This allowed the successful assembly of chromosome-level haplotypes for the tetraploid species sugarcane (*Saccharum spontaneum*)[37], alfalfa (*Medicago sativa*)[38,39], potato[40], rhubarb (*Rheum officinale*)[41] and Jerusalem artichoke (*Helianthus tuberosus*)[42].

Given the lack of diploid reference genomes for some species, HapHiC[43] was developed, based on ALLHiC, to permit the differentiation of specific haplotypes directly through the use of Hi-C data, without the need of a reference sequence for the diploid genome. However, many autopolyploid species have mixed reproduction systems that combine sexual and asexual reproduction, resulting in alleles that may be either highly divergent or very similar[21,43,44], as exemplified in potato. By combining Hi-C with highly accurate long-read sequencing, such as High-Fidelity (HiFi) reads, it is possible to distinguish and assemble certain genomic regions with substantial sequence differences, but it is impractical to obtain a complete haplotype-resolved full chromosome. Additional genomic information, such as that obtained through the sequencing of individual pollen grains[21], selfing progeny populations at low coverage[44] and F_1 offspring[45], is needed to achieve fully haplotype-resolved assemblies. Ultra-long reads are also essential for resolving complex genomic regions, and these have contributed to the successful assembly of telomere-to-telomere chromosome-level contigs in potato[46]. The interspecific hybridization and back-crossing shaped complex polyploid genomes of the sugarcane cultivars 'R570' (ref. 5) and 'ZZ1' (ref. 47), which show variable preferential pairing among chromosomes, thus requiring multiple sources of genetic information and various sequencing technologies. The recently developed Pore-C[48] sequencing technology, which combines nanopore sequencing with chromatin conformation capture, allows the identification of long-range interactions between genomic regions. This approach showed promise for phasing of a haplotype-resolved assembly for the tetraploid rose cultivar *Rosa hybrida* 'Samantha'[49]. Clearly, high-accuracy and ultra-long sequencing data greatly assist in the assembly of complex polyploid genomes. How to use Hi-C or genetic map information more effectively is the next frontier for the development of new assembly algorithms.

Although genome assemblies have progressed rapidly, gene annotation of the resulting assembled genomes remains underde- veloped[50]. Even in model species such as human (*Homo sapiens*) and *Arabidopsis thaliana*, catalogues of functional splicing variants and non-coding elements are far from complete[51]. State-of-the-art annotation pipelines integrate multiple layers of evidence, including transcriptome sequencing (RNA sequencing) data from various tissues, alignments to related proteins in other species and ab initio gene prediction. Annotating polyploid genomes is thus more challenging than annotating diploid genomes because the former harbour more duplications and divergence between haplotypes, which introduce mapping biases and limit the accuracy of hidden Markov model-based gene prediction methods. Mapping biases result from reduced unique alignments across all haplotypes or false positive mapping of homologues when aligning to a single representative haplotype[52], whereas ab initio prediction is problematic in highly divergent regions owing to inadequate initial training sets[50]. For example, a Benchmarking Universal Single-Copy Orthologs (BUSCO) analysis showed an average of 97% completeness in the 4 haplotypes (genome sequences) ofan autotetraploid cultivated potato but only 92% completeness in the gene annotation (coding sequences), suggesting that up to 5% of genes may be missing from the annotation[21]. These discrepancies mainly arise from a lack of haplotype-aware annotation, which is more problematic in incompletely phased assemblies. Furthermore, single-molecule RNA-sequencing technologies, which can generate full-length RNA and cDNA sequences, may also contribute to gene prediction accuracy and diminish mapping noise[53,54]. In rapeseed, the number of nucleotide-binding leucine-rich repeat receptor genes increased from 345 to 715 following manual curation from resistance gene enrichment sequencing and other homology-based evidence[55]. In recent years, deep learning tools[56,57] have also shown a lot of potential for increasing accuracy. The application of models trained with tree-of-life datasets[58,59], with the ability to fine-tune for species-specific or accession-specific contexts, may lead to substantial breakthroughs in genome annotation over the next decade.

Genetic variants in polyploid crop genomes

Polyploid crops, characterized by multiple sets of chromosomes, may carry more genetic variants than diploid crops, contributing to their adaptability and resilience[60]. Although the paired chromosomes of autopolyploid species originated from a whole-genome duplication (WGD) event (Fig. 2a), their haplotypes still harbour substantial differences, particularly in terms of structural variations. For example, in the autotetraploid potato cultivar 'Otava', each constituent haplotype contains 11.0–42.5% unique sequences that could not be aligned to the other haplotypes[21]. The low proportion of shared sequences between haplotypes suggests sequence loss and neofunctionalization of redundant sequences[61]. In hexaploid wheat, on average, about 30% of homoeolog triads (A, B and D copies) show biased expression, with some of these associated with epigenetic modifications and transposable element variations[62,63]. Hence, the genetic redundancy associated with a polyploid genome provides a raw resource for the generation of genetic and epigenetic variation, thereby shaping the diverse variation landscape of polyploids.

Two additional features are present in polyploids: double reduction (DR) and homoeologous exchange (HE) (Fig. 2b). In autopolyploids, homologous chromosomes may pair with each other. This results in a complex pairing structure, such as multivalent pairing and crossover during the first meiotic division in an autopolyploid, thereby yielding a single gamete that receives two identical segments of the same chromosome. This phenomenon, referred to as a DR event, is commonly observed in potato[64] and alfalfa[65]. DR events often increase the frequency of homozygous alleles and complicate the prediction of genome structure. In allopolyploids, HE refers to the phenomenon in which chromosomes from each ancestral species may occasionally pair and engage in crossover during meiosis. Rapeseed[66] and bread wheat[67] have been extensively studied in the context of HE, revealing that HE in wheat is primarily regulated by *Pairing homoeologous gene 1* (*Ph1*); disabling *Ph1* promotes the pairing of homoeologous chromosomes[68]. Mispairing between ancestrally related chromosomes can result in the exchange of large DNA fragments between homoeologous chromosomes, boosting genetic diversity and occasionally inducing structural variations[66], which disrupts genome organization and usually has negative effects on traits such as fertility and heading time[68].

Polyploid genomes also provide a buffer against deleterious variants, as exemplified by the faster accumulation of mutation load in allopolyploid cotton compared with its diploid counterpart[69]. Clonal propagation may contribute to the accumulation of deleterious mutations, as many deleterious mutations remain in a heterozygous state within the autotetraploid potato genome[70]. To maintain genetic variation to favour adaptation under changing conditions, vegetative reproduction is more frequent in autopolyploid species than in diploid species[8,71]. In turn, the limited recombination during such asexual reproduction contributes to the accumulation of genetic variants in diverse haplotypes, shaping the extensive variation landscape of autopolyploid genomes (Fig. 2c). By contrast, variations in allopolyploids are primarily driven by the hybridization of ancestral subgenomes (Fig. 2d). There is limited evidence for direct comparisons between haplotypes within subgenomes. As allopolyploids are generally not obligate outcrossers, most studies assume these haplotypes to be homozygous, which may partly reflect the focus of current research on broader subgenome-level analyses rather than on intra-subgenomic variation[6,72,73].

Table 1 | Sequencing and assembly strategies for polyploid genomes

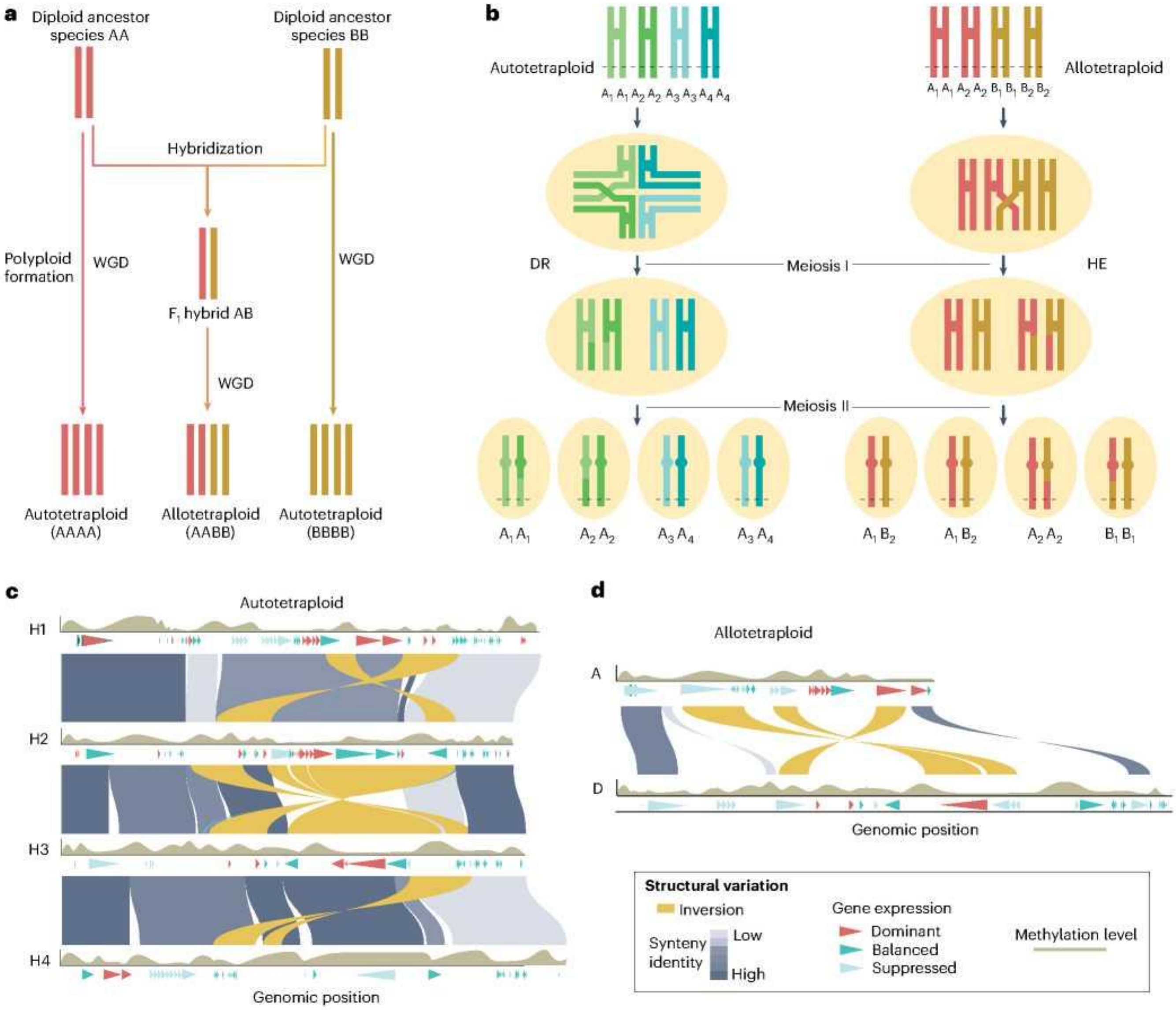
Type of polyploid	Species	Latin name	Karyotype	Initial assembly	T2T	Phasing and/or partitioned strategy	Phasing quality	
							Accuracy	Completeness
Auto-allopolyploid	Sugarcane[5] (cv. R570)	*Saccharum* spp.	$12x \approx 114$	HiFi	x	Genetic map, single-chromosome sorting Hi-C and BioNano scaffolding	+++	+++
	Sugarcane[47] (cv. ZZ1)	*Saccharum* spp.	$12x \approx 114$	HiFi	x	Hi-C with ALLHiC	+++	++
Autopolyploid	Sugarcane[37] (AP85-441)	*Saccharum spontaneum*	$4x = 40$	CLR	x	Hi-C with ALLHiC	++	++
	Potato[21] (cv. Otava)	*Solanum tuberosum*	$4x = 48$	HiFi	x	Single-pollen sequencing Hi-C	+++	+++
	Potato[44] (cv. C88)	*S. tuberosum*	$4x = 48$	HiFi	x	S_1 population genetic map, polyploid graph binning and Hi-C	+++	+++
	Potato[45] (cv. Altus)	*S. tuberosum*	$4x = 48$	HiFi	x	*k*-mer-based genetic map from F_1 population	+++	++
	Potato[46] (cv. C88)	*S. tuberosum*	$4x = 48$	HiFi	x	hifiasm HiFi and ONT hybrid assembly polyploid graph binning	+++	++
Allopolyploid	Wheat	*T. aestivum*	$6x = 42$	HiFi + ONT[168,169]; HiFi[170]	√	Progenitor and/ or phylogenomics based[1 ,25], *k*-mer based[171], repeat based[31,172]	/	
	Cotton Rapeseed	*G. hirsutum* *B. napus*	$4x = 52$ $4x = 38$	HiFi + ONT[173,174]; HiFi[73]; CLR[175] HiFi + ONT[176]; HiFi[177]; CLR[120]; ONT[85]	√ √			
	Teff[28]	*Eragrostis tef*	$4x = 20$	CLR	x	LTR based		
	Papaver	*Papaver setigerum*	$4x = 44$	HiFi + ONT[178]; ONT[179]	x	Satellite based		

Note: 'x' indicates assemblies that are not telomere-to-telomere (T2T); '√' indicates T2T assemblies; '+++' indicates highly accurate or complete phasing results; '++' indicates good phasing;

'+' indicates lower accuracy or completeness; '/' indicates genome assemblies that do not focus on phasing of homologous chromosomes. CLR, continuous long reads; ONT, Oxford Nanopore Technologies; LTR, long terminal repeat.

Fig. 2 | Establishment and variation landscape in polyploid genomes.

a, Diagram of the pathways leading to the formation of polyploid species. Taking tetraploids as an example: an autotetraploid (AAAA or BBBB) can be created through a WGD event of a diploid ancestor[181] or through the combination of unreduced 2*n* gametes (AA or BB) from diploid ancestors[182]. For a diploid F_1 hybrid offspring (AB) derived from different species, homologous chromosomes are likely to have difficulty pairing properly during meiosis. Following WGD events, normal meiosis can occur, resulting in the formation of allotetraploids (AABB)[183]. **b**, Diagram showing two notable features of polyploid genomes, namely, DR and HE. Dashed lines indicate example loci at specific positions on the chromosomes. Following a DR event, the two identical alleles on the sister chromatids enter the same gamete after meiosis (A_1A_1 and A_2A_2). Following an HE event, alleles from the same ancestor or the same subgenome in an allopolyploid enter the same gamete after meiosis (A_2A_2 and B_1B_1). DR and HE events increase the genetic diversity of gametes and add complexity to the genetic mechanisms of polyploid genomes. **c**,**d**, Diagrams showing synteny and the genomic variation landscape in polyploids, as exemplified by typical observations in autotetraploids (**c**) and allotetraploids (**d**). In autotetraploids, genome sequences are commonly referred to as haplotypes (for example, H1, H2, H3 and H4)[21,44], whereas in allotetraploids, they are called subgenomes (for example, AADD)[184,185]. Autopolyploids primarily show genetic variation between haplotypes, whereas allopolyploids are characterized by pronounced divergence between subgenomes originating from distinct progenitor species. The diagram highlights syntenic regions (grey), SVs, differential methylation levels (brown) and differentially expressed alleles (triangles).



Pan-genomes of polyploid crops

The pan-genome refers to the collection of all genome sequences in a population, which has less reference bias than individual genomes and provides a more accurate and complete representation of global genomic variation within the population[74–76]. Since the release of the first plant pan-genome for soybean (*Glycine max*)[77], pan-genome research for diploid plant species has rapidly expanded to rice (*Oryza sativa*)[78], maize (*Zea mays*)[79] and barley (*Hordeum vulgare*)[80,81]. Similar to discoveries in diploid pan-genomes, polyploid pan-genome studies have also focused on core and variable gene sets across individuals, and the proportion of these gene sets may vary between species. In studies of polyploid genomes, 66%

of genes were classified as core genes (present in all individual genomes) in the wheat pan-genome, while only 42% were classified as core genes in broomcorn millet (*Panicum miliaceum* L.)[82]. Sometimes, polyploid and diploid genomes can be combined into one pan-genome analysis, as shown in a recent study[73] of cotton, providing insight into the convergence and divergence of tetraploid and diploid species. Graph-based pan-genome[83] is another useful approach for capturing a wide representation of genetic variants[84]. This approach has contributed to the identification of variants in resequenced populations, revealing the domestication centre of broomcorn millet[82] and structural variants (SVs) associated with key traits in rapeseed[85].

Currently, only a few pan-genomes for polyploid plant species are available[86], limiting our understanding of polyploid species. The complexity of polyploid genomes raises challenges in sequence alignment, graph construction and variation detection. Therefore, the analytical approaches or tools normally used to study diploid genomes may need to be updated accordingly for polyploid genomes. For example, regarding the treatment of divergent genome sequences between polyploids, should each genome sequence be treated separately, or should the full set of genome sequences be considered in an individual? The answer may vary depending on the research purpose. Furthermore, whereas genotyping diploids from low-depth resequencing (based on graph-based pan-genomes) has been widely explored, similar improvements are still needed for polyploids.

Genetic analysis of polyploid genomes

Population genetics analysis of polyploid species

After initial polyploidization, especially during the first formation of polyploid species, multiple ploidy levels can coexist within natural populations. Interploidy gene flow frequently occurs between diploid and polyploid individuals through the formation of 2*n* gametes and WGD events[87,88]. As a result, polyploid genomes often show elevated levels of genetic diversity, driven by introgression between ploidy levels. However, this diversity can be reduced by post-WGD bottlenecks and subsequent transitions to selfing, as observed in *Capsella bursa-pastoris*[89]. Phylogenetic analysis and comparison of genomic sequences have allowed the identification of numerous introgressed segments in the genomes of allopolyploid species, such as hexaploid wheat[90,91] and tetraploid rapeseed[92], and in autopolyploids, such as tetraploid potato[93], alfalfa[94] and *Arabidopsis arenosa*[95]. The presence of introgressed segments from diploid potato species into the genome of tetraploid potato results in an even smaller genetic distance between cultivated tetraploid potatoes and wild diploid relatives compared with that between tetraploid potatoes and their diploid progenitors[93]. The genome of hexaploid bread wheat (AABBDD) contains as much as 4–32% genomic sequences originating from introgressions from tetraploid wheat populations (AABB), whereas the genome of goatgrass (*Aegilops speltoides*, DD) contributed minimal gene flow to the D subgenome of bread wheat[96]. This imbalance between subgenome origins results in different levels of genetic diversity among the A, B and D subgenomes. However, not all polyploid species readily undergo interploidy introgression. In hexaploid oat (*Avena sativa* L.), for example, the structural chromosomal differences among the genomes of diploid, tetraploid and hexaploid species can lead to abnormal chromosome pairings during interploidy hybridization, posing a challenge in breeding efforts[97]. In the origin and domestication history of these important economic crops, the loci carried by segments introgressed from wild species have played a crucial part in enhancing adaptability to environmental factors such as day length, temperature and expanding cultivation range. As has been extensively studied in rapeseed, frequent introgression has contributed to phenotypic diversification and adaptability across various morphological types[92].

The subgenomes of allopolyploid species do not always have the same architectures. Differential subgenomic selection during domestication and cultivation history has facilitated the development of diverse traits in cultivated species. In tetraploid rapeseed populations (AACC), owing to breeding barriers between *B. napus* and the C genome species, frequent crossing from the A species results in greater nucleotide diversity in subgenome A than in subgenome C. The selection imposed on the A subgenome of rapeseed may have promoted stress resistance and greater oil accumulation, whereas selection imposed on the C subgenome may have contributed to the improvement of winter, semi-winter and spring accessions[98].

The domestication history of cultivated cotton (AADD) offers another example of differential contribution from individual subgenomes to desirable traits in a cultivated crop species. From 620 pairs of homoeologous genes subjected to domestication selection, 428 (70%) were located on the D subgenome, indicating strong asymmetric selection in the genome of cultivated cotton[99]. In upland cotton cultivars, the gene *FORMIN HOMOLOGY INTERACTING PROTEIN 1* (*FIP1*), predicted to be associated with fibre elongation, showed domestication selection signals in the A subgenome but not in its D homoeolog; two other genes, *4-COUMARATE:COA LIGASE* (*4CL*) and *CHALCONE SYNTHASE* (*CHS*), thought to be responsible for the transition of cotton fibre colour from brown to white, were domesticated only in the D subgenome.

The asymmetric accumulation of dominant alleles may be driven by selection. Polyploid genomes can accumulate more deleterious mutations than diploid genomes when one functional homoeolog remains, a phenomenon referred to as homoeologous masking. In addition to functional genes being subjected to different directions of selection in subgenomes during domestication, the accumulation of deleterious mutations also exerts asymmetrical effects within subgenomes. A comparison of mutation sites across multiple cotton genomes revealed that genomic regions in the D subgenome had a higher mutation rate or a more relaxed background selection than that in the A subgenome, suggesting that homoeologous masking may act unequally between subgenomes[12].

Linkage mapping based on genetic population

Genetic linkage analysis uses segregating populations derived from crosses between parents with contrasting target traits and represents a common and effective method to reveal genotype–phenotype relationships. For this purpose, many single nucleotide polymorphism (SNP) arrays have been developed to rapidly genotype these segregating populations, aiming to construct genetic maps and assist in selecting desired offspring. For hexaploid bread wheat, SNP chips are available with 90,000 (ref. 100), 660,000 (ref. 101) and 820,000 (ref. 102) SNP molecular markers. For tetraploid potato, the Potato Array, developed in 2012 by Felcher et al.[103] and consisting of 8,303 markers, is still widely used. For autotetraploid sugarcane, a 100,000 SNP array has been an efficient genetic tool in genotyping and genetic mapping[104]. Whole-genome sequencing of parental lines and their derived segregating populations, compared with existing SNP arrays, can help identify molecular markers specific to those target populations, enabling the construction of accurate genetic maps for gene mapping. In allopolyploid plant species, the different subgenomes behave relatively independently, and genetic segregation follows the same rules as in diploid species. Thus, genetic analysis tools initially developed for diploid species, such as IciMapping[105] and QTLMap (https://github.com/ofilangi/qtlmap), have been widely used for self-fertilizing and outcrossing populations of allopolyploid species, such as rapeseed, bread wheat and cotton, enabling the mapping of agronomic traits.

In some cases, phenotypes may be associated with genetic mechanisms specific to polyploid species, making it difficult to accurately map a given trait to linked markers. For example, in allopolyploid species, HEs may lead to the formation of unbalanced gametes, reduced fertility and selective disadvantage, thereby affecting phenotypic traits[106,107]. HE can cause SVs within chromosomes or gene bodies, potentially contributing to new quantitative trait loci (QTLs) or functional genes[108]. The QTL–HE colocalized markers show skewed marker segregation patterns, which may result in the exclusion of these markers during mapping. Therefore, when performing a linkage analysis of species with widespread HE, attention needs to be given to markers in deleted or duplicated regions. For autopolyploid species, genetic linkage analysis faces more challenges than in allopolyploid and diploid species due to the underlying complexity of possible meiotic configurations: (1) as homologous chromosomes can freely pair, each zygote of an autopolyploid species can present 36 possible combinations, with 400 combinations for a hexaploid species, both of which are much greater than the 4 classical zygote combinations for hybrids between diploid species; (2) DR can generate new combinations of alleles and lead to higher frequencies of the involved alleles in gametes[109]—for example, in the progeny of a tetraploid potato with the GGGg genotype, 1 individual derived from gg gametes, expected due to DR, had a dwarf phenotype[110]; and (3) during meiosis, homologous chromosomes can form multivalent structures, leading to a higher frequency of recombination between them[111]. In some autopolyploid species, preferential pairing between homologous chromosomes can also occur, showing chromosome pairing behaviours similar to those seen in allopolyploid species, as reported in tetraploid roses[112]. Extreme preferential pairing can affect marker selection during genetic mapping.

The development of genetic analysis tools for polyploid species has primarily focused on more accurate identification and genotyping of markers, overcoming technical false positives and false negatives arising during genome sequencing and read alignment. Many tools suitable for polyploid species have been developed, such as

GBSpoly genotyping[113], MAPpoly[114], QTLpoly (https://github.com/ guilherme-pereira/QTLpoly), TetraploidSNPMap[115] and polyploid QTL-seq[116]. Since 2020, researchers have gradually incorporated the frequency of genetic events, such as multivalent pairing and DR, into their tools and models, as illustrated by QTL mapping in a segregating tetraploid potato population[117]. This technical advance will undoubtedly provide better tools for the study of polyploid species.

Association mapping

Genome-wide association studies (GWAS) have been widely used in allopolyploid species, such as rapeseed, cotton, peanut and bread wheat, to identify genetic loci that influence plant growth, gene expression and metabolite levels[118,119]. In addition to using SNP markers, obtained through genome resequencing of a diversity panel for the species of interest, GWAS have also been conducted in cotton and rapeseed using SV genotypes derived from pan-genome techniques, identifying candidate loci that were not previously detected by SNP-based GWAS[120,121]. Subgenomes and chromosomes are typically processed independently in GWAS, with the alleles being treated as separate markers. However, alleles at loci within each subgenome are not independent entities, as interactions between subgenomes and epistasis can influence trait expression and should be considered[122]. Recently, studies in bread wheat have revealed that combining multiple homoeologous loci as a single genetic unit in GWAS detected more loci influencing phenotypes, thereby affording stronger explanatory power for phenotypic variation, compared with SNP-based GWAS[123]. Clearly, more attention should be given to research based on genotype combinations that have the potential to uncover unique forms of genetic regulation specific to polyploid genomes.

GWAS has also been applied in autopolyploid species, such as potato[124] and sweet potato[125], but progress has been relatively limited, contributing little to trait elucidation. The implementation of GWAS to autopolyploid plant species lags behind that in allopolyploid species, mainly due to the following reasons: (1) the limited number of sequenced populations; (2) scarcity of polymorphic markers, primarily relying on SNP-based arrays; (3) lack of available tools, with GWASpoly being the main option[126]; and (4) the presence of several alleles in the genome, complicating fine mapping after GWAS. The existing GWAS workflows were primarily established for diploid species and are not applicable to autopolyploid genomes, which may be the main bottleneck. This lag is inconsistent with the critical role of autopolyploid species in achieving food security, highlighting the urgent need for more research to focus on this area.

Intricacies for variant calling and dissection of genetic architecture

Variant calling is a fundamental step in genetic mapping and population analysis. The technological advancements and challenges of variant calling have been extensively studied and discussed in human and model animal genomes through benchmark dataset analysis[52]. However, the associated challenges in the context of plant genomes have received limited examination. The presence of numerous repetitive sequences and frequent WGD events in plant genomes makes it challenging to accurately map sequencing reads and identify genomic variants (Fig. 3). In addition, plant populations have higher genetic diversity than human populations, such that using a single reference genome to call variants may induce reference bias and decrease the reliability of read mapping. Hardigan et al.[93] and Huang et al.[127] conducted analyses on the same diploid potato genome sequencing data using different pipelines and found that this resulted in 3- to 4-fold differences in the number of SNPs detected in each study (46.8 million versus 10.5 million in 10 diploid wild species; 26.6 million versus 8.1 million in 10 diploid landraces). These differences underscore the fragility of variant detection in plant genomes and the potential risks of applying analytical pipelines designed for human genomes to plant genomes. Importantly, the quality of the reference genome is key for any downstream analysis. It may be beneficial to consider replacing all short-read-based reference genomes with new and improved assemblies based on long reads to enhance the accuracy of read mapping. The continued development of true pan-genome graph-based mapping will further improve this situation.

This conundrum is even more pronounced with polyploid genomes due to the large number of possible genotypes, decreasing the quality of the results (Fig. 3). Cooke et al.[128] conducted experiments simulating variant calling on polyploid genomes using a mixture of multiple datasets for human genome sequences. The study showed that, at a sequencing depth of 30×, the error rates of SNP calling for tetraploid and hexaploid genomes were 18× and 33× higher, respectively, than that for diploid genomes, with false negative rates 8× and 16× higher, respectively, than that for diploid genomes. Experiments on autotriploid banana (*Musa* sp.) datasets showed that only 38% of genotype calls were consistent among different datasets and tools[128]. This study highlights the need for caution when interpreting variant calling output in polyploid genome analyses. It is crucial to establish filtering criteria based on quality controls of the data and the segregation principles of the population being analysed and to perform cross-validation between tools. It is worth noting that using filtering criteria based on researchers' experience can lead to inconsistencies or bias in the results across different research projects. Currently, sequence alignment and variant calling, based on haplotype-aware assembly and pan-genome approaches, can yield better results[129,130], particularly in regions that differ substantially from the reference genome. Such regions constitute a large proportion of polyploid plant genomes. As more polyploid plant genomes are sequenced and studied, it will be crucial to establish benchmark datasets for polyploid genomes to guide the development of bioinformatics tools and optimize experimental workflows.

Genomics-enabled breeding of polyploid crops

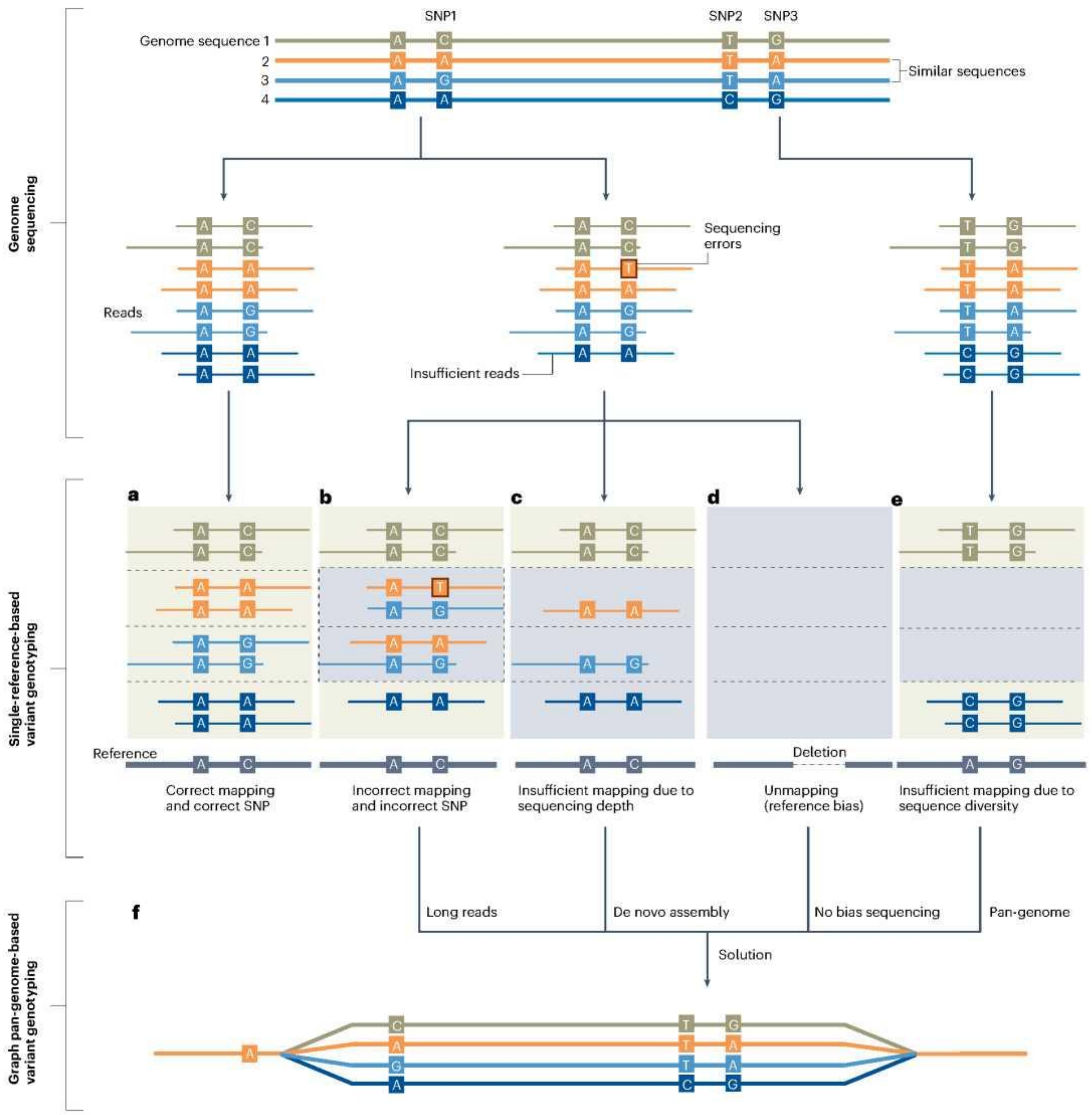
Domestication and genomic selection of polyploid crops

Crop domestication involves the selection of key traits, such as the non-shattering of seeds after maturity[131], larger fruits[132,133] and lower levels of toxic substances in edible plant tissues[134,135]. In subsequent breeding and improvement, various beneficial agricultural traits gradually become concentrated in a few domesticated individuals, together with a loss of genetic diversity within the overall population. The domestication of most polyploid species has required multiple rounds of hybridization and introgression of genomic segments to accumulate agronomic traits present in one or a few individuals. Modern sugar- canes are derived from interspecific crosses between *S. officinarum*, known for its thick stalks and high sugar content, and *S. spontaneum*, valued for its disease resistance and ratooning capacity[46]. Hexaploid bread wheat was derived from the hybridization of wild emmer wheat (*Triticum turgidum* ssp. *dicoccoides*)[136], which carries mutations that prevent seed shattering, and diploid *Aegilops tauschii*, which harbours a gene repertoire for environmental adaptation[137].

After the formation of polyploid species, or their initial domestication, subsequent stages of domestication and selection can lead to the development of subspecies, or accessions, with distinct agronomic traits. For example, cultivated tetraploid peanut (AABB) formed two subspecies (*A. hypogaea* subsp. *hypogaea* and *A. hypogaea* subsp. *fastigiata*) after domestication in different environments[138]; it also resulted in populations with distinct agronomic traits (such as small pod and large pod) in different cultivation centres[119]. Thus, after poly- ploidization a genome retains the capacity for artificial domestication and improvement, suggesting that polyploid crops have great potential for the development of diverse traits. In contrast to allopolyploid crops, autopolyploid crops have not undergone hybridization between distant ancestral species. Instead, they can increase their genetic diversity through genome doubling and introgression from wild species, thereby improving agronomic traits such as environmental adaptability and stress resistance[93]. Genomic analysis can help clarify ancient domestication events and efforts of artificial selection, and elucidate how wild species have contributed to the traits of modern cultivated crops. These capabilities allow us to imagine new ways to exploit untapped genetic resources and accelerate genetic gain through modern breeding methods.

Genomic selection, combined with genomic prediction, uses genomic information to predict phenotypic traits and has become an important technology in crop and livestock breeding for achieving rapid and precise selection of superior genotypes[139,140]. With the increasing numbers of sequenced populations and the application of machine learning in this field, genomic selection has led to substantial progress in polyploid crop breeding in recent years[141,142]. For example, the prediction of genetic values using 4 statistical models (Bayesian LASSO, ridge regression, reproducing kernel Hilbert space and partial least squares regression) on 167 sugarcane clones for 10 traits showed approximately equivalent accuracy[143]. The phenotypes of a training set consisting of 1,334 elite wheat breeding lines were evaluated over 3 field seasons and used to show the potential of genomic selection in wheat breeding[144]. In tetraploid potato, 384 chipping clones (for potatoes destined for the chip industry) were used for genomic selection to accelerate breeding for key traits, including chip colour, chip quality, specific gravity, vine maturity (foliar senescence) and total yield[145]. It is worth noting that the application of genomic selection to polyploid species may be influenced by factors such as multiple alleles and gene dosage effects[141], which can make phenotypic predictions less accurate than in diploid species. Machine learning models tailored for polyploid species and populations hold promise for alleviating this issue.

Fig. 3 | Challenges in detecting variation in polyploid genomes. Diagram showing the limitations of using a haploid-based genome as a reference for variant detection of polyploid genomes. Taking SNPs (which can also be replaced by SVs; for simplicity, we use SNPs here) in a region of a polyploid genome as an example: genome sequences 1, 2, 3 and 4 may represent either haplotypes in an autopolyploid or subgenomes in an allopolyploid. Three SNPs (SNP1, SNP2 and SNP3) are present across the 4 genome sequences, with sequences 2 and 3 being more similar to each other. **a**, Ideally, all genome sequences in the polyploid are covered by sufficient reads of adequate length to cover and support the detection of variants. **b**–**e**, In a single-reference-based variant detection pipeline, incorrect SNP genotypes may occur due to sequencing errors or incorrect mapping from similar sequences (**b**), insufficient mapping from low sequencing depth (**c**), absence of corresponding sequences in the reference genome (**d**), or substantial discrepancies between the query and the reference sequences (**e**). **f**, A graph-based pan-genome constructed from the assembly and alignment of the four haplotypes can capture variant information over longer distances, facilitating the detection of accurate SNP genotypes.



Genome editing in breeding of polyploid plant species

For trait improvement in traditional breeding methods, valuable genomic fragments from wild germplasm are introduced into a cultivated crop through hybridization, or new genetic variants are induced through mutation breeding. The multiple gene sets in the genomes of polyploid species create functional redundancy, making it challenging to improve traits through simultaneous targeting of these genes. Moreover, heterozygous plants generated by hybridization or mutation breeding require lengthy and labour-intensive back-crossing to achieve homozygosity at target loci.

Gene editing technologies relying on zinc-finger nucleases, transcription activator-like effector nucleases (TALENs) or clustered regularly interspaced short palindromic repeats (CRISPR) can generate targeted mutations in the genome—including small deletions, replacements and inversions—and can target multiple genes simultaneously, making these approaches highly valuable in polyploid breeding. In 2014, TALEN-based technology was used to simultaneously knock out the *Mildew resistance locus O* (*MLO*) genes in the A, B and D subgenomes of hexaploid bread wheat, achieving broad-spectrum resistance to powdery mildew and setting a successful precedent for gene editing in polyploid species[146]. Subsequently, gene editing technologies have been successfully applied to the breeding of polyploid species, including potato[147], rapeseed[148], sugarcane[149], strawberry[150], banana[151] and switchgrass (*Panicum virgatum* L.)[152], to quickly improve the vigour, yield, biotic stress resistance, abiotic stress tolerance and herbicide resistance of cultivated varieties.

In addition to editing the genes that directly determine agronomic traits, gene editing can enhance breeding techniques by modifying fertility-related traits. Editing the dominant gene *Ms2* that confers complete male sterility restored the fertility of male gametes from male-sterile wheat lines, facilitating the establishment of new fertile varieties with excellent traits[153]. By editing functional genes, breeders can induce haploid formation or create desirable recombination events, thereby shortening the time required to accumulate multiple beneficial alleles in a single individual, which has been extensively reviewed[154]. The ability of CRISPR-mediated gene editing to target multiple alleles simultaneously in a polyploid genome has been shown in high-ploidy crops, such as tetraploid and hexaploid wheat[155] and switchgrass[152], allowing the production of homozygous individuals in the T_0 generation. However, to reach this outcome, it is necessary to fully consider the conserved regions of multiple gene sets when designing the single guide RNAs[156,157]. The considerable variation in the numbers and sequences of genes or alleles presents challenges in designing highly efficient single guide RNAs for polyploid genomes, and further efforts are needed to develop appropriate tools.

Synthetic breeding of polyploid crops

Further enhancing crop yield or introducing traits for resilience to extreme climatic conditions strongly depends on the exploration of diversity and the utilization of wild germplasms. The synthetic reconstitution of hexaploid wheat represents an important achievement that exemplifies the use of wild germplasm resources. Hexaploid bread wheat was formed through the natural hybridization and chromosome doubling of tetraploid durum wheat (*T. turgidum*, AABB) and diploid wild goatgrass (*A. tauschii*, DD)[158]. Thanks to mild reproductive barriers between these two species, breeders were able to generate multiple synthesized hexaploid wheat lines by selecting various AABB tetraploid and DD diploid combinations (Fig. 4a). Subsequently, favourable traits from wild relatives can be transferred to modern wheat varieties via hybridization and artificial selection. The synthetic reconstitution of hexaploid wheat thus allows the introduction of previously untapped 'wild' genetic resources across different chromosomal regions. This approach has been used by the International Maize and Wheat Improvement Center (CIMMYT), which has widely distributed the resulting materials in equatorial growing zones, resulting in many derived lines and new varieties[159].

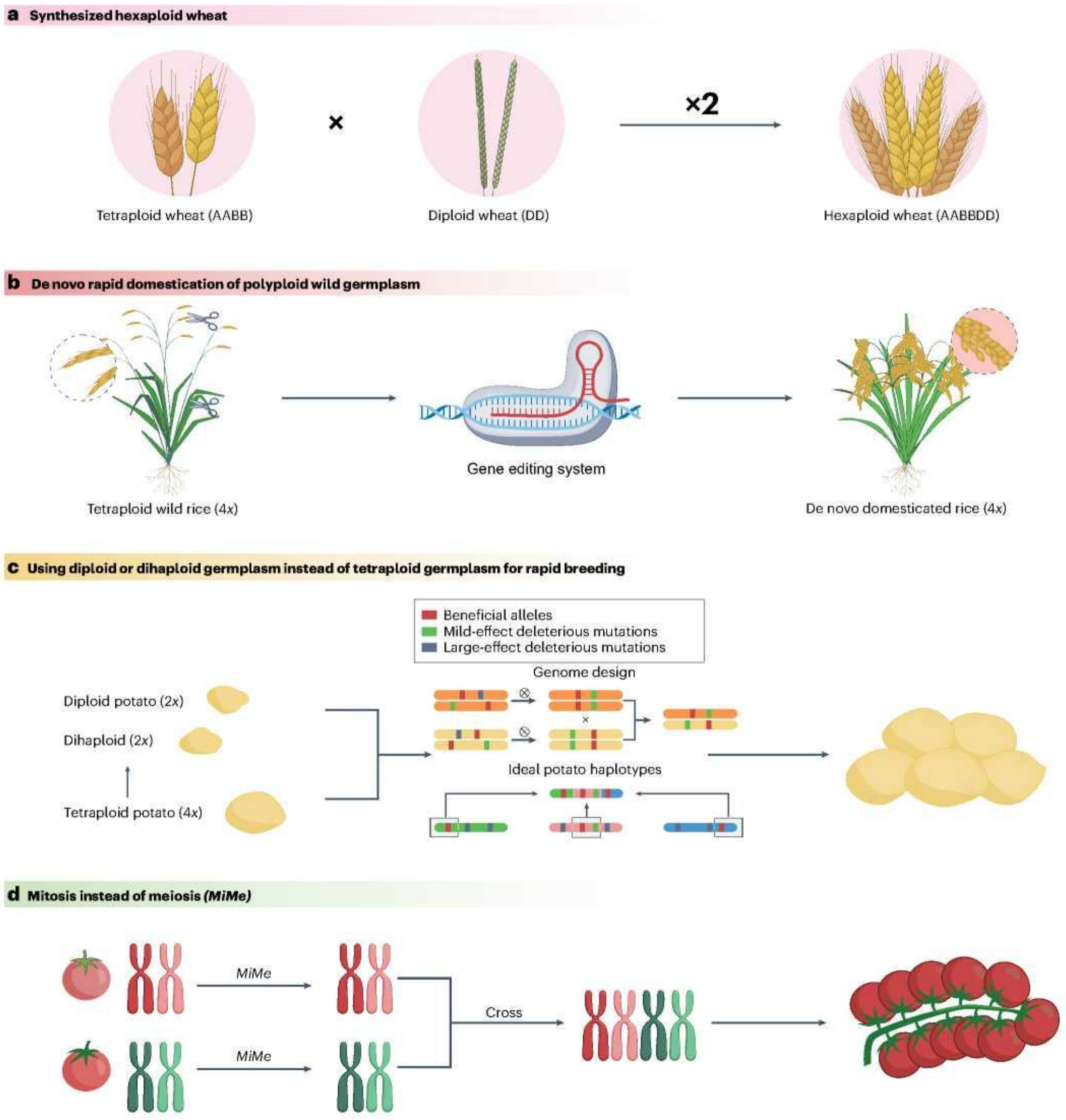
Breeding by hybridization and artificial selection to generate polyploid crops has traditionally been a lengthy procedure. In recent years, guided by synthetic biology principles and using innovative breeding techniques such as gene editing, direct modification of the genomes of existing wild polyploid species for rapid domestication has been a promising approach (Fig. 4b). In 2021, Yu et al.[160] selected the crop-like tetraploid wild rice *Oryza alta* and developed a genetic transformation system. They then edited known conserved rice domestication genes in the *O. alta* genome, resulting in domesticated phenotypes for traits such as plant height and seed characteristics[160]. This study showed that wild species can be rapidly domesticated using modern biotechnological methods by editing targeted traits, which provides new avenues for accelerating the use of available diversity.

The combinations ofalleles in autopolyploid species are complex, making it challenging to achieve homozygosity within target regions during hybrid breeding as rapidly as in diploid hybrid breeding. The potato hybrid breeding programme offers an innovative approach to molecular design breeding for polyploid crops (Fig. 4c). Modern commercial potatoes are autotetraploids, whereas many wild species and landraces with similar agronomic traits to those of cultivated potatoes (and hence breeding value) are diploids[161]. Gene editing has been used to break self-incompatibility in diploid potatoes[162], thereby paving the way for the genetic mapping of major deleterious mutations across the diploid genome[163]. Ultimately, genome-assisted selection was used to bypass deleterious mutations, allowing the successful construction of nearly homozygous inbred parent lines and heterozygous F_1 off- spring[164]. The F_1 diploid potatoes produced through this method have considerable commercial value. To harness the advantages of hybrid progeny, in terms of yield and adaptability, for polyploid crops[165], diploid potatoes will need to undergo further polyploid genome design and breeding.

Another potentially useful plant breeding strategy is the mitosis instead of meiosis (*MiMe*) system, which can be used to produce clonal and unreduced gametes that can later be used for hybridization to generate polyploid offspring[166] (Fig. 4d). Wang et al.[167] established the *MiMe* system in tomato, obtaining diploid gametes from three diploid tomato hybrids and conducting self-pollination and hybridization experiments. The resulting tetraploid self-fertilized offspring showed strong genotypic and phenotypic similarity to their diploid parents. The F_1 tetraploid offspring from this hybridization retained four sets of chromosomes from the parents and showed normal plant and fruit phenotypes while retaining the full complement of disease-resistance loci. This example of breeding diploid to tetraploid tomatoes illustrates a potential breeding strategy for polyploid species through gene editing, enabling a full utilization of heterosis in polyploid crops. The larger polyploid genome also provides additional capacity for introducing more advantageous variations to meet the needs of future breeding programmes.

For centuries, the genetic improvement of polyploid crops has generally lagged behind that of diploid crops, especially for asexually generated crops, such as sugarcane and potatoes. The complex gamete combinations, along with inbreeding depression caused by deleterious mutations, have hindered the effective application of conventional hybrid breeding techniques to these species. Currently, scientists are exploring revolutionary new approaches, such as rapid domestication ofwild polyploids and the use of diploids in place of tetraploid germplasms, with very encouraging outcomes. These innovative breeding concepts can be implemented in various crops with the assistance of technologies such as genome editing, genomic selection, genetic transformation and gene function prediction, breaking through the long-standing limitations of polyploid breeding.

Fig. 4 | Genomic breeding for polyploid crops. a, Synthetic hexaploid wheat generated by hybridizing tetraploid wheat with diploid grass wheat[158,159]. **b**, De novo rapid domestication of tetraploid wild rice through a gene editing system[160]. **c**, Genome design strategy for developing diploid hybrid potato through diploid or dihaploid potato from tetraploid potato[164]. The label '×' within a circle indicates selfing, and the '×' without a circle indicates crossing. **d**, The *MiMe* system produces unreduced gametes, which can later be used for hybridization to generate polyploid offspring in tomato[167]



Conclusion and outlook

The advancement of sequencing technologies has made genome assembly of many species possible. However, the number of polyploid species for which high-quality reference genomes exist is still limited, even though polyploid species comprise nearly 50% of both plant species in general and those responsible for agricultural production. Apart from the key allopolyploid crops—wheat, cotton and rapeseed—with well-defined ancestral genomes, most polyploid plant species have few assembled species. Research in areas such as population genetics and functional genomics is also lacking for polyploid species.

This challenge arises from the structure of polyploid genomes, which impose limitations in the assembly pipeline. Recently, researchers achieved promising results using Pore-C technology for scaffolding genomic fragments for the genome assembly of tetraploid rose[49]. Transformative advancements through the combination of new algorithms and sequencing technologies are enhancing research on reference genomes, generating telomere-to-telomere genomes, pan-genomes and graph-based pan-genomes in more polyploid species.

Breeding methods based on hybridization and phenotypic selection are relatively inefficient in polyploid species.

Advancements in synthetic biology have provided insight and tools, leading researchers to explore innovative breeding strategies, including de novo domestication of wild species[160] and the de novo design of ideal crops[161]. Progress in the following aspects will greatly facilitate the success of strategies for synthetic breeding: (1) conducting multi-omics studies on a wider array of wild populations to identify additional genetic resources for breeding; (2) integrating various types of variation, especially haplotype-specific markers, into genomic selection, while expanding the application of artificial intelligence to design ideal genomes; and (3) expanding techniques in areas such as gene editing and the regulation of chromosome recombination, which are essential for obtaining the desired genomes in target species.

The development of agronomic traits in polyploid species is undoubtedly a complex biological question. Our Review provides a broader perspective, using the most extensively studied polyploids as examples. These findings are also applicable to a wide range of polyploid studies to facilitate polyploid research. We expect that in the coming decades, biotechnological advancements will lead to breakthroughs in each breeding stage, enabling the rapid accumulation of beneficial traits in complex genomes.

References

1. Grant, V. *Plant Speciation* (Columbia Univ. Press, 1981).
2. Masterson, J. Stomatal size in fossil plants: evidence for polyploidy in majority of angiosperms. *Science* **264**, 421–424 (1994).
3. Cui, L. et al. Widespread genome duplications throughout the history of flowering plants. *Genome Res.* **16**, 738–749 (2006).
4. Dinesh Babu, K. S. et al. A short review on sugarcane: its domestication, molecular manipulations and future perspectives. *Genet. Resour. Crop Evol.* **69**, 2623–2643 (2022).
5. Healey, A. L. et al. The complex polyploid genome architecture of sugarcane. *Nature* **628**, 804–810 (2024).
6. The International Wheat Genome Sequencing Consortium (IWGSC). A chromosome-based draft sequence of the hexaploid bread wheat (*Triticum aestivum*) genome. *Science* **345**, 1251788 (2014).
7. Stokstad, E. The new potato. *Science* **363**, 574–577 (2019).
8. Sattler, M. C., Carvalho, C. R. & Clarindo, W. R. The polyploidy and its key role in plant breeding. *Planta* **243**, 281–296 (2016).
9. Bird, K. A., VanBuren, R., Puzey, J. R. & Edger, P. P. The causes and consequences of subgenome dominance in hybrids and recent polyploids. *N. Phytol.* **220**, 87–93 (2018).
10. Woodhouse, M. R. et al. Origin, inheritance, and gene regulatory consequences of genome dominance in polyploids. *Proc. Natl Acad. Sci. USA* **111**, 5283–5288 (2014).
11. Lian, Q. et al. Acquisition of deleterious mutations during potato polyploidization. *J. Integr. Plant Biol.* **61**, 7–11 (2019).
12. Conover, J. L. & Wendel, J. F. Deleterious mutations accumulate faster in allopolyploid than diploid cotton (*Gossypium*) and unequally between subgenomes. *Mol. Biol. Evol.* **39**, msac024 (2022).
13. Stebbins, G. L. in *Advances in Genetics* Vol. 1 (ed. Demerec, M.) 403–429 (Elsevier, 1947).
14. Spoelhof, J. P., Soltis, P. S. & Soltis, D. E. Pure polyploidy: closing the gaps in autopolyploid research. *J. Syst. Evol.* **55**, 340–352 (2017).
15. Song, Y. et al. Phased gap-free genome assembly of octoploid cultivated strawberry illustrates the genetic and epigenetic divergence among subgenomes. *Hortic. Res.* **11**, uhad252 (2024).
16. Jin, X. et al. Haplotype-resolved genomes of wild octoploid progenitors illuminate genomic diversifications from wild relatives to cultivated strawberry. *Nat. Plants* **9**, 1252–1266 (2023).
17. Chalhoub, B. et al. Early allopolyploid evolution in the post-Neolithic *Brassica napus* oilseed genome. *Science* **345**, 950–953 (2014).
18. Li, F. et al. Genome sequence of cultivated Upland cotton (*Gossypium hirsutum* TM-1) provides insights into genome evolution. *Nat. Biotechnol.* **33**, 524–530 (2015).
19. Cheng, H., Concepcion, G. T., Feng, X., Zhang, H. & Li, H. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nat. Methods* **18**, 170–175 (2021).
20. Nurk, S. et al. HiCanu: accurate assembly of segmental duplications, satellites, and allelic variants from high-fidelity long reads. *Genome Res.* **30**, 1291–1305 (2020).

21. Sun, H. et al. Chromosome-scale and haplotype-resolved genome assembly of a tetraploid potato cultivar. *Nat. Genet.* **54**, 342–348 (2022).
22. Hoopes, G. et al. Phased, chromosome-scale genome assemblies of tetraploid potato reveal a complex genome, transcriptome, and predicted proteome landscape underpinning genetic diversity. *Mol. Plant* **15**, 520–536 (2022).
23. Edger, P. P. et al. Origin and evolution of the octoploid strawberry genome. *Nat. Genet.* **51**, 541–547 (2019).
24. Zheng, Y. et al. Allele-aware chromosome-scale assembly of the allopolyploid genome of hexaploid Ma bamboo (*Dendrocalamus latiflorus* Munro). *J. Integr. Plant Biol.* **64**, 649–670 (2022).
25. Marcussen, T. et al. Ancient hybridizations among the ancestral genomes of bread wheat. *Science* **345**, 1250092 (2014).
26. Zhang, T. et al. Sequencing of allotetraploid cotton (*Gossypium hirsutum* L. acc. TM-1) provides a resource for fiber improvement. *Nat. Biotechnol.* **33**, 531–537 (2015).
27. Bertioli, D. J. et al. The genome sequence of segmental allotetraploid peanut *Arachis hypogaea*. *Nat. Genet.* **51**, 877–884 (2019).
28. VanBuren, R. et al. Exceptional subgenome stability and functional divergence in the allotetraploid Ethiopian cereal teff. *Nat. Commun.* **11**, 884 (2020).
29. Miao, J. et al. Chromosome-scale assembly and analysis of biomass crop *Miscanthus lutarioriparius* genome. *Nat. Commun.* **12**, 2458 (2021).
30. Mitros, T. et al. Genome biology of the paleotetraploid perennial biomass crop *Miscanthus*. *Nat. Commun.* **11**, 5442 (2020).
31. Session, A. M. & Rokhsar, D. S. Transposon signatures of allopolyploid genome evolution. *Nat. Commun.* **14**, 3180 (2023).
32. Chen, H. et al. A chromosome-scale genome assembly of *Artemisia argyi* reveals unbiased subgenome evolution and key contributions of gene duplication to volatile terpenoid diversity. *Plant Commun.* **4**, 100516 (2023).
33. Miao, Y. et al. Genome sequencing reveals chromosome fusion and extensive expansion of genes related to secondary metabolism in *Artemisia argyi*. *Plant Biotechnol. J.* **20**, 1902–1915 (2022).
34. Song, A. et al. Analyses of a chromosome-scale genome assembly reveal the origin and evolution of cultivated chrysanthemum. *Nat. Commun.* **14**, 2021 (2023).
35. Yang, J. et al. Haplotype-resolved sweet potato genome traces back its hexaploidization history. *Nat. Plants* **3**, 696–703 (2017).
36. Zhang, X., Zhang, S., Zhao, Q., Ming, R. & Tang, H. Assembly of allele-aware, chromosomal-scale autopolyploid genomes based on Hi-C data. *Nat. Plants* **5**, 833–845 (2019).
37. Zhang, J. et al. Allele-defined genome of the autopolyploid sugarcane *Saccharum spontaneum* L. *Nat. Genet.* **50**, 1565–1573 (2018).
38. Chen, H. et al. Allele-aware chromosome-level genome assembly and efficient transgene-free genome editing for the autotetraploid cultivated alfalfa. *Nat. Commun.* **11**, 2494 (2020).
39. Long, R. et al. Genome assembly of alfalfa cultivar Zhongmu-4 and identification of SNPs associated with agronomic traits. *Genom. Proteom. Bioinform.* **20**, 14–28 (2022).
40. Wang, F. et al. The autotetraploid potato genome provides insights into highly heterozygous species. *Plant Biotechnol. J.* **20**, 1996–2005 (2022).
41. Zhang, H. et al. The haplotype-resolved genome assembly of autotetraploid rhubarb *Rheum officinale* provides insights into its genome evolution and massive accumulation of anthraquinones. *Plant Commun.* **5**, 100677 (2024).
42. Wang, S. et al. Haplotype-resolved chromosome-level genome of hexaploid Jerusalem artichoke provides insights into its origin, evolution, and inulin metabolism. *Plant Commun.* **5**, 100767 (2024).
43. Zeng, X. et al. Chromosome-level scaffolding of haplotype-resolved assemblies using Hi-C data without reference genomes. *Nat. Plants* **10**, 1184–1120 (2024).
44. Bao, Z. et al. Genome architecture and tetrasomic inheritance of autotetraploid potato. *Mol. Plant* **15**, 1211–1226 (2022).
45. Serra Mari, R. et al. Haplotype-resolved assembly of a tetraploid potato genome using long reads and low-depth offspring data. *Genome Biol.* **25**, 26 (2024).
46. Cheng, H., Asri, M., Lucas, J., Koren, S. & Li, H. Scalable telomere-to-telomere assembly for diploid and polyploid genomes with double graph. *Nat. Methods* **21**, 967–970 (2024).

47. Bao, Y. et al. A chromosomal-scale genome assembly of modern cultivated hybrid sugarcane provides insights into origination and evolution. *Nat. Commun.* **15**, 3041 (2024).
48. Deshpande, A. S. et al. Identifying synergistic high-order 3D chromatin conformations from genome-scale nanopore concatemer sequencing. *Nat. Biotechnol.* **40**, 1488–1499 (2022).
49. Zhang, Z. et al. Haplotype-resolved genome assembly and resequencing provide insights into the origin and breeding of modern rose. *Nat. Plants* **10**, 1659–1671 (2024).
50. Gabriel, L. et al. BRAKER3: fully automated genome annotation using RNA-seq and protein evidence with GeneMark-ETP, AUGUSTUS, and TSEBRA. *Genome Res.* **34**, 769–777 (2024).
51. Zhang, R. et al. A high-resolution single-molecule sequencing-based *Arabidopsis* transcriptome using novel methods of Iso-seq analysis. *Genome Biol.* **23**, 149 (2022).
52. Olson, N. D. et al. Variant calling and benchmarking in an era of complete human genome sequences. *Nat. Rev. Genet.* **24**, 464–483 (2023).
53. Hook, P. W. & Timp, W. Beyond assembly: the increasing flexibility of single-molecule sequencing technology. *Nat. Rev. Genet.* **24**, 627–641 (2023).
54. Wang, B. et al. Variant phasing and haplotypic expression from long-read sequencing in maize. *Commun. Biol.* **3**, 78 (2020).
55. Wu, J., Mukhopadhyay, S. & Pérez-López Sr, E. Resistance gene enrichment sequencing refines the *Brassica napus* NLRome. *Plant Physiol.* **197**, kiae631 (2025).
56. Holst, F. et al. Helixer—de novo prediction of primary eukaryotic gene models combining deep learning and a hidden Markov model. Preprint at *bioRxiv* https://doi.org/10.1101/ 2023.02.06.527280 (2023).
57. Gabriel, L., Becker, F., Hoff, K. J. & Stanke, M. Tiberius: end-to-end deep learning with an HMM for gene prediction. *Bioinformatics* **40**, btae685 (2024).
58. Zhai, J. et al. Cross-species modeling of plant genomes at single nucleotide resolution using a pre-trained DNA language model. *Proc. Natl Acad. Sci. USA* **122**, e2421738122 (2025).
59. Benegas, G., Batra, S. S. & Song, Y. S. DNA language models are powerful predictors of genome-wide variant effects. *Proc. Natl Acad. Sci. USA* **120**, e2311219120 (2023).
60. Comai, L. The advantages and disadvantages of being polyploid. *Nat. Rev. Genet.* **6**, 836–846 (2005).
61. Cheng, F. et al. Gene retention, fractionation and subgenome differences in polyploid plants. *Nat. Plants* **4**, 258–268 (2018).
62. Yuan, J. et al. Dynamic and reversible DNA methylation changes induced by genome separation and merger of polyploid wheat. *BMC Biol.* **18**, 171 (2020).
63. Ramírez-González, R. H. et al. The transcriptional landscape of polyploid wheat. *Science* **361**, eaar6089 (2018).
64. Bourke, P. M., Voorrips, R. E., Visser, R. G. & Maliepaard, C. The double-reduction landscape in tetraploid potato as revealed by a high-density linkage map. *Genetics* **201**, 853–863 (2015).
65. Julier, B. et al. Construction of two genetic linkage maps in cultivated tetraploid alfalfa (*Medicago sativa*) using microsatellite and AFLP markers. *BMC Plant Biol.* **3**, 9 (2003).
66. Hurgobin, B. et al. Homoeologous exchange is a major cause of gene presence/absence variation in the amphidiploid *Brassica napus*. *Plant Biotechnol. J.* **16**, 1265–1274 (2018).
67. Zhang, Z. et al. Homoeologous exchanges occur through intragenic recombination generating novel transcripts and proteins in wheat and other polyploids. *Proc. Natl Acad. Sci. USA* **117**, 14561–14571 (2020).
68. Fan, C. et al. The KL system in wheat permits homoeologous crossing over between closely related chromosomes. *Crop J.* **11**, 808–816 (2023).
69. Hu, G. et al. Evolutionary dynamics of chromatin structure and duplicate gene expression in diploid and allopolyploid cotton. *Mol. Biol. Evol.* **41**, msae095 (2024).
70. Wu, Y. et al. Phylogenomic discovery of deleterious mutations facilitates hybrid potato breeding. *Cell* **186**, 2313–2328.e15 (2023).
71. Herben, T., Suda, J. & Klimešová, J. Polyploid species rely on vegetative reproduction more than diploids: a re-examination of the old hypothesis. *Ann. Bot.* **120**, 341–349 (2017).
72. Walkowiak, S. et al. Multiple wheat genomes reveal global variation in modern breeding. *Nature* **588**, 277–283 (2020).

73. Li, J. et al. Convergence and divergence of diploid and tetraploid cotton genomes. *Nat. Genet.* **56**, 2562–2573 (2024).
74. Hu, H. et al. Plant pangenomics, current practice and future direction. *Agric. Commun.* **2**, 100039 (2024).
75. Schreiber, M., Jayakodi, M., Stein, N. & Mascher, M. Plant pangenomes for crop improvement, biodiversity and evolution. *Nat. Rev. Genet.* **25**, 563–577 (2024).
76. Shi, J., Tian, Z., Lai, J. & Huang, X. Plant pan-genomics and its applications. *Mol. Plant* **16**, 168–186 (2023).
77. Li, Y. et al. De novo assembly of soybean wild relatives for pan-genome analysis of diversity and agronomic traits. *Nat. Biotechnol.* **32**, 1045–1052 (2014).
78. Zhao, Q. et al. Pan-genome analysis highlights the extent of genomic variation in cultivated and wild rice. *Nat. Genet.* **50**, 278–284 (2018).
79. Hirsch, C. N. et al. Insights into the maize pan-genome and pan-transcriptome. *Plant Cell* **26**, 121–135 (2014).
80. Jayakodi, M. et al. The barley pan-genome reveals the hidden legacy of mutation breeding. *Nature* **588**, 284–289 (2020).
81. Jayakodi, M. et al. Structural variation in the pangenome of wild and domesticated barley. *Nature* **636**, 654–662 (2024).
82. Chen, J. et al. Pangenome analysis reveals genomic variations associated with domestication traits in broomcorn millet. *Nat. Genet.* **55**, 2243–2254 (2023).
83. Hickey, G. et al. Genotyping structural variants in pangenome graphs using the vg toolkit. *Genome Biol.* **21**, 35 (2020).
84. MacNish, T. R. et al. *Brassica* Panache: a multi-species graph pangenome representing presence absence variation across forty-one *Brassica* genomes. *Plant Genome* **18**, e20535 (2025).
85. Zhang, Y. et al. Structural variation reshapes population gene expression and trait variation in 2,105 *Brassica napus* accessions. *Nat. Genet.* **56**, 2538–2550 (2024).
86. Sun, H. et al. The phased pan-genome of tetraploid European potato. *Nature* **642**, 389–397 (2025).
87. Nieto Feliner, G., Casacuberta, J. & Wendel, J. F. Genomics of evolutionary novelty in hybrids and polyploids. *Front. Genet.* **11**, 792 (2020).
88. Brown, M. R., Abbott, R. J. & Twyford, A. D. The emerging importance of cross-ploidy hybridisation and introgression. *Mol. Ecol.* **33**, e17315 (2024).
89. Douglas, G. M. et al. Hybrid origins and the earliest stages of diploidization in the highly successful recent polyploid *Capsella bursa-pastoris*. *Proc. Natl Acad. Sci. USA* **112**, 2806–2811 (2015).
90. Wang, Z. et al. Dispersed emergence and protracted domestication of polyploid wheat uncovered by mosaic ancestral haploblock inference. *Nat. Commun.* **13**, 3891 (2022).
91. Zhao, X. et al. Population genomics unravels the Holocene history of bread wheat and its relatives. *Nat. Plants* **9**, 403–419 (2023).
92. Wang, T. et al. Interploidy introgression shaped adaptation during the origin and domestication history of *Brassica napus*. *Mol. Biol. Evol.* **40**, msad199 (2023).
93. Hardigan, M. A. et al. Genome diversity of tuber-bearing *Solanum* uncovers complex evolutionary history and targets of domestication in the cultivated potato. *Proc. Natl Acad. Sci. USA* **114**, E9999 (2017).
94. Shen, C. et al. The chromosome-level genome sequence of the autotetraploid alfalfa and resequencing of core germplasms provide genomic resources for alfalfa research. *Mol. Plant* **13**, 1250–1261 (2020).
95. Monnahan, P. et al. Pervasive population genomic consequences of genome duplication in *Arabidopsis arenosa*. *Nat. Ecol. Evol.* **3**, 457–468 (2019).
96. Zhou, Y. et al. Triticum population sequencing provides insights into wheat adaptation. *Nat. Genet.* **52**, 1412–1422 (2020).
97. Kamal, N. et al. The mosaic oat genome gives insights into a uniquely healthy cereal crop. *Nature* **606**, 113–119 (2022).
98. Lu, K. et al. Whole-genome resequencing reveals *Brassica napus* origin and genetic loci involved in its improvement. *Nat. Commun.* **10**, 1154 (2019).
99. Wang, M. et al. Asymmetric subgenome selection and *cis*-regulatory divergence during cotton domestication. *Nat. Genet.* **49**, 579–587 (2017).
100. Wang, S. et al. Characterization of polyploid wheat genomic diversity using a high-density 90,000 single nucleotide polymorphism array. *Plant Biotechnol. J.* **12**, 787–796 (2014).
101. Sun, C. et al. The Wheat 660K SNP array demonstrates great potential for marker-assisted selection in polyploid wheat. *Plant*

Biotechnol. J. **18**, 1354–1360 (2020).

102. Winfield, M. O. et al. High-density SNP genotyping array for hexaploid wheat and its secondary and tertiary gene pool. *Plant Biotechnol. J.* **14**, 1195–1206 (2016).
103. Felcher, K. J. et al. Integration of two diploid potato linkage maps with the potato genome sequence. *PLoS ONE* **7**, e36347 (2012).
104. You, Q. et al. Development of an Axiom Sugarcane100K SNP array for genetic map construction and QTL identification. *Theor. Appl. Genet.* **132**, 2829–2845 (2019).
105. Meng, L., Li, H., Zhang, L. & Wang, J. QTL IciMapping: integrated software for genetic linkage map construction and quantitative trait locus mapping in biparental populations. *Crop J.* **3**, 269–283 (2015).
106. Deb, S. K., Edger, P. P., Pires, J. C. & McKain, M. R. Patterns, mechanisms, and consequences of homoeologous exchange in allopolyploid angiosperms: a genomic and epigenomic perspective. *N. Phytol.* **238**, 2284–2304 (2023).
107. Wu, Y. et al. Genomic mosaicism due to homoeologous exchange generates extensive phenotypic diversity in nascent allopolyploids. *Natl Sci. Rev.* **8**, nwaa277 (2021).
108. Stein, A. et al. Mapping of homoeologous chromosome exchanges influencing quantitative trait variation in *Brassica napus*. *Plant Biotechnol. J.* **15**, 1478–1489 (2017).
109. Jiang, L., Ren, X. & Wu, R. Computational characterization of double reduction in autotetraploid natural populations. *Plant J.* **105**, 1703–1709 (2021).
110. Manrique-Carpintero, N. C. et al. Genome reduction in tetraploid potato reveals genetic load, haplotype variation, and loci associated with agronomic traits. *Front. Plant Sci.* **9**, 944 (2018).
111. Otto, S. P. The evolutionary consequences of polyploidy. *Cell* **131**, 452–462 (2007).
112. Bourke, P. M. et al. Partial preferential chromosome pairing is genotype dependent in tetraploid rose. *Plant J.* **90**, 330–343 (2017).
113. Wadl, P. A. et al. Genetic diversity and population structure of the USDA sweetpotato (*Ipomoea batatas*) germplasm collections using GBSpoly. *Front. Plant Sci.* **9**, 1166 (2018).
114. Mollinari, M. et al. Unraveling the hexaploid sweetpotato inheritance using ultra-dense multilocus mapping. *G3* **10**, 281–292 (2020).
115. Hackett, C. A., Boskamp, B., Vogogias, A., Preedy, K. F. & Milne, I. TetraploidSNPMap: software for linkage analysis and QTL mapping in autotetraploid populations using SNP dosage data. *J. Hered.* **108**, 438–442 (2017).
116. Takagi, H. et al. QTL -seq: rapid mapping of quantitative trait loci in rice by whole genome resequencing of DNA from two bulked populations. *Plant J.* **74**, 174–183 (2013).
117. Chen, J. et al. A tetrasomic inheritance model and likelihood- based method for mapping quantitative trait loci in autotetraploid species. *N. Phytol.* **230**, 387–398 (2021).
118. Ma, Z. et al. Resequencing a core collection of upland cotton identifies genomic variation and loci influencing fiber quality and yield. *Nat. Genet.* **50**, 803–813 (2018).
119. Lu, Q. et al. A genomic variation map provides insights into peanut diversity in China and associations with 28 agronomic traits. *Nat. Genet.* **56**, 530–540 (2024).
120. Song, J.-M. et al. Eight high-quality genomes reveal pan-genome architecture and ecotype differentiation of *Brassica napus*. *Nat. Plants* **6**, 34–45 (2020).
121. Jin, S. et al. Structural variation (SV)-based pan-genome and GWAS reveal the impacts of SVs on the speciation and diversification of allotetraploid cottons. *Mol. Plant* **16**, 678–693 (2023).
122. Dwivedi, S. L., Heslop-Harrison, P., Amas, J., Ortiz, R. & Edwards, D. Epistasis and pleiotropy-induced variation for plant breeding. *Plant Biotechnol. J.* **22**, 2788–2807 (2024).
123. Jia, J. et al. Genome resources for the elite bread wheat cultivar Aikang 58 and mining of elite homeologous haplotypes for accelerating wheat improvement. *Mol. Plant* **16**, 1893–1910 (2023).
124. Sharma, S. K. et al. Linkage disequilibrium and evaluation of genome-wide association mapping models in tetraploid potato. *G3* **8**, 3185–3202 (2018).
125. Haque, E. et al. Polyploid GWAS reveals the basis of molecular marker development for complex breeding traits including

starch content in the storage roots of sweet potato. *Front. Plant Sci.* **14**, 1181909 (2023).

126. Chen, M. et al. Genome-wide identification of agronomically important genes in outcrossing crops using OutcrossSeq. *Mol. Plant* **14**, 556–570 (2021).
127. Huang, B., Spooner, D. M. & Liang, Q. Genome diversity of the potato. *Proc. Natl Acad. Sci. USA* **115**, E6392 (2018).
128. Cooke, D. P., Wedge, D. C. & Lunter, G. Benchmarking small-variant genotyping in polyploids. *Genome Res.* **32**, 403–408 (2022).
129. Liao, W.-W. et al. A draft human pangenome reference. *Nature* **617**, 312–324 (2023).
130. Groza, C. et al. Pangenome graphs improve the analysis of structural variants in rare genetic diseases. *Nat. Commun.* **15**, 657 (2024).
131. Dong, Y. & Wang, Y.-Z. Seed shattering: from models to crops. *Front. Plant Sci.* **6**, 476 (2015).
132. Van Dam, J., Kooman, P. L. & Struik, P. C. Effects of temperature and photoperiod on early growth and final number of tubers in potato (*Solanum tuberosum* L.). *Potato Res.* **39**, 51–62 (1996).
133. Martínez-García, J. F., García-Martínez, J. L., Bou, J. & Prat, S. The interaction of gibberellins and photoperiod in the control of potato tuberization. *J. Plant Growth Regul.* **20**, 377–386 (2001).
134. Johns, T. Detoxification function of geophagy and domestication of the potato. *J. Chem. Ecol.* **12**, 635–646 (1986).
135. Shang, Y. et al. Biosynthesis, regulation, and domestication of bitterness in cucumber. *Science* **346**, 1084–1088 (2014).
136. Avni, R. et al. Wild emmer genome architecture and diversity elucidate wheat evolution and domestication. *Science* **357**, 93–97 (2017).
137. Jia, J. et al. *Aegilops tauschii* draft genome sequence reveals a gene repertoire for wheat adaptation. *Nature* **496**, 91–95 (2013).
138. Zheng, Z. et al. Chloroplast and whole-genome sequencing shed light on the evolutionary history and phenotypic diversification of peanuts. *Nat. Genet.* **56**, 1975–1984 (2024).
139. Desta, Z. A. & Ortiz, R. Genomic selection: genome-wide prediction in plant improvement. *Trends Plant Sci.* **19**, 592–601 (2014).
140. Meuwissen, T., Hayes, B. & Goddard, M. Accelerating improvement of livestock with genomic selection. *Annu. Rev. Anim. Biosci.* **1**, 221–237 (2013).
141. Batista, L. G., Mello, V. H., Souza, A. P. & Margarido, G. R. Genomic prediction with allele dosage information in highly polyploid species. *Theor. Appl. Genet.* **135**, 723–739 (2022).
142. Yadav, S. et al. Accelerating genetic gain in sugarcane breeding using genomic selection. *Agronomy* **10**, 585 (2020).
143. Gouy, M. et al. Experimental assessment of the accuracy of genomic selection in sugarcane. *Theor. Appl. Genet.* **126**, 2575–2586 (2013).
144. Bonnett, D. et al. Response to early generation genomic selection for yield in wheat. *Front. Plant Sci.* **12**, 718611 (2022).
145. Pandey, J., Scheuring, D. C., Koym, J. W., Endelman, J. B. & Vales, M. I. Genomic selection and genome-wide association studies in tetraploid chipping potatoes. *Plant Genome* **16**, e20297 (2023).
146. Wang, Y. et al. Simultaneous editing of three homoeoalleles in hexaploid bread wheat confers heritable resistance to powdery mildew. *Nat. Biotechnol.* **32**, 947–951 (2014).
147. Johansen, I. E. et al. High efficacy full allelic CRISPR/Cas9 gene editing in tetraploid potato. *Sci. Rep.* **9**, 17715 (2019).
148. Li, J., Yu, X., Zhang, C., Li, N. & Zhao, J. The application of CRISPR/ Cas technologies to *Brassica* crops: current progress and future perspectives. *aBIOTECH* **3**, 146–161 (2022).
149. Mall, A. K. et al. CRISPR/Cas-mediated genome editing for sugarcane improvement. *Sugar Tech.* **27**, 1–13 (2024).
150. Vondracek, K., Altpeter, F., Liu, T. & Lee, S. Advances in genomics and genome editing for improving strawberry (*Fragaria × ananassa*). *Front. Genet.* **15**, 1382445 (2024).
151. Tripathi, L., Ntui, V. O. & Tripathi, J. N. CRISPR/Cas9-based genome editing of banana for disease resistance. *Curr. Opin. Plant Biol.* **56**, 118–126 (2020).
152. Liu, Y. et al. Targeted mutagenesis in tetraploid switchgrass (*Panicum virgatum* L.) using CRISPR/Cas9. *Plant Biotechnol. J.* **16**, 381–393 (2018).
153. Tang, H. et al. Fertility recovery of wheat male sterility controlled by Ms2 using CRISPR/Cas9. *Plant Biotechnol. J.* **19**, 224–226

(2021).

154. Zhu, H., Li, C. & Gao, C. Applications of CRISPR–Cas in agriculture and plant biotechnology. *Nat. Rev. Mol. Cell Biol.* **21**, 661–677 (2020).
155. Zhang, Y. et al. Efficient and transgene-free genome editing in wheat through transient expression of CRISPR/Cas9 DNA or RNA. *Nat. Commun.* **7**, 12617 (2016).
156. Li, J., Li, Y. & Ma, L. Recent advances in CRISPR/Cas9 and applications for wheat functional genomics and breeding. *aBIOTECH* **2**, 375–385 (2021).
157. Zaman, Q. U., Li, C., Cheng, H. & Hu, Q. Genome editing opens a new era of genetic improvement in polyploid crops. *Crop J.* **7**, 141–150 (2019).
158. Levy, A. A. & Feldman, M. Evolution and origin of bread wheat. *Plant Cell* **34**, 2549–2567 (2022).
159. Rosyara, U. et al. Genetic contribution of synthetic hexaploid wheat to CIMMYT's spring bread wheat breeding germplasm. *Sci. Rep.* **9**, 12355 (2019).
160. Yu, H. et al. A route to de novo domestication of wild allotetraploid rice. *Cell* **5**, 1156–1170 (2021).
161. Cheng, L. et al. Leveraging a phased pangenome for haplotype design of hybrid potato. *Nature* **640**, 408–417 (2025).
162. Ye, M. et al. Generation of self-compatible diploid potato by knockout of *S-RNase*. *Nat. Plants* **4**, 651–654 (2018).
163. Zhang, C. et al. The genetic basis of inbreeding depression in potato. *Nat. Genet.* **51**, 374–378 (2019).
164. Zhang, C. et al. Genome design of hybrid potato. *Cell* **184**, 3873–3883 (2021).
165. Washburn, J. D. & Birchler, J. A. Polyploids as a "model system" for the study of heterosis. *Plant Reprod.* **27**, 1–5 (2014).
166. Mieulet, D. et al. Turning rice meiosis into mitosis. *Cell Res.* **26**, 1242–1254 (2016).
167. Wang, Y. et al. Harnessing clonal gametes in hybrid crops to engineer polyploid genomes. *Nat. Genet.* **56**, 1075–1079 (2024).
168. Liu, S. et al. A telomere-to-telomere genome assembly coupled with multi-omic data provides insights into the evolution of hexaploid bread wheat. *Nat. Genet.* **57**, 1008–1020 (2025).
169. Wang, Z. et al. Near-complete assembly and comprehensive annotation of the wheat Chinese Spring genome. *Mol. Plant* **18**, 892–907 (2025).
170. Athiyannan, N. et al. Long-read genome sequencing of bread wheat facilitates disease resistance gene cloning. *Nat. Genet.* **54**, 227–231 (2022).
171. Jia, K.-H. et al. SubPhaser: a robust allopolyploid subgenome phasing method based on subgenome-specific k-mers. *N. Phytol.* **235**, 801–809 (2022).
172. Lovell, J. T. et al. Genomic mechanisms of climate adaptation in polyploid bioenergy switchgrass. *Nature* **590**, 438–444 (2021).
173. Yan, H. et al. Post-polyploidization centromere evolution in cotton. *Nat. Genet.* **57**, 1021–1030 (2025).
174. Hu, G. et al. A telomere-to-telomere genome assembly of cotton provides insights into centromere evolution and short-season adaptation. *Nat. Genet.* **57**, 1031–1043 (2025).
175. Sreedasyam, A. et al. Genome resources for three modern cotton lines guide future breeding efforts. *Nat. Plants* **10**, 1039–1051 (2024).
176. Li, B. et al. A gap-free reference genome reveals structural variations associated with flowering time in rapeseed (*Brassica napus*). *Hortic. Res.* **10**, uhad171 (2023).
177. Zhao, C. et al. Chromosome-scale genome assembly-assisted identification of *Brassica napus BnDCPA1* for improvement of plant architecture and yield heterosis. *Plant Commun.* **5**, 100854 (2024).
178. Gao, S. et al. The centromere landscapes of four karyotypically diverse *Papaver* species provide insights into chromosome evolution and speciation. *Cell Genom.* **4**, 100626 (2024).
179. Yang, X. et al. Three chromosome-scale *Papaver* genomes reveal punctuated patchwork evolution of the morphinan and noscapine biosynthesis pathway. *Nat. Commun.* **12**, 6030 (2021).
180. Kumar, S. et al. TimeTree 5: an expanded resource for species divergence times. *Mol. Biol. Evol.* **39**, msac174 (2022).
181. Clark, J. W. & Donoghue, P. C. Whole-genome duplication and plant macroevolution. *Trends Plant Sci.* **23**, 933–945 (2018).
182. Kreiner, J. M., Kron, P. & Husband, B. C. Evolutionary dynamics of unreduced gametes. *Trends Genet.* **33**, 583–593 (2017).
183. Ramsey, J. & Schemske, D. W. Pathways, mechanisms, and rates of polyploid formation in flowering plants. *Annu. Rev. Ecol. Syst.* **29**, 467–501 (1998).

184. Chen, Z. J. et al. Genomic diversifications of five *Gossypium* allopolyploid species and their impact on cotton improvement. *Nat. Genet.* **52**, 525–533 (2020).
185. Dai, D. et al. Unzipped chromosome-level genomes reveal allopolyploid nematode origin pattern as unreduced gamete hybridization. *Nat. Commun.* **14**, 7156 (2023).

Acknowledgements

The assembled polyploid genomes in Fig. 1 are based on our own compilation from publicly available resources, primarily the PubPlant database, which may not include all reported species. We thank the developers and curators of PubPlant for their efforts in collecting and maintaining the data. We thank W. J. Lucas, J. Zhang, Z. Liu and N. Wang for project discussions and critical comments. This project was funded by a Guangdong Major Project of Basic and Applied Basic Research (2021B0301030004), the National Natural Science Foundation of China (grant numbers 32488302 and 31991181), the Fundamental Research Fund for the Central Universities (77000-12240018), the Agricultural Science and Technology Innovation Program (CAAS-ZDRW202404) and a Shenzhen Outstanding Talents Training Fund. Z.B. is supported by the grants to D. Weigel.

Author contributions

Q.Z., Z.B. and S.H. co-designed the framework of the paper. L.C., Z.B. and Q.Z. wrote the draft of the paper and generated the figures. Q.K., L.L., N.S. and S.H. contributed to writing and revision of the paper.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41477-025-02088-5.