




Evolution of Plant Genome Size and Composition

Bing He ^{1,#}, Wanfei Liu ^{1,#}, Jianyang Li ^{1,2}, Siwei Xiong ¹, Jing Jia ^{1,3}, Qiang Lin ^{1,*},
Hailin Liu ^{1,*}, Peng Cui ^{1,*}

¹Guangdong Laboratory for Lingnan Modern Agriculture, Genome Analysis Laboratory of the Ministry of Agriculture and Rural Area, Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen 518120, China

²College of Agronomy, Qingdao Agricultural University, Qingdao 266109, China

³State Key Laboratory of Crop Stress Adaptation and Improvement, School of Life Sciences, Henan University, Kaifeng 475004, China

*Corresponding authors: linqiang@caas.cn (Lin Q), liuhailin@caas.cn (Liu H), cuipeng@caas.cn (Cui P).

#Equal contribution.

Handling Editor: Xiangfeng Wang

Abstract

The rapid development of sequencing technology has led to an explosion of plant genome data, opening up more opportunities for research in the field of comparative evolutionary analysis of plant genomes. In this review, we focus on changes in plant genome size and composition, examining the effects of polyploidy, whole-genome duplication, and alternations in transposable elements on plant genome architecture and evolution, respectively. In addition, to address gaps in the available information, we also collected and analyzed 234 representative plant genome data as a supplement. We aim to provide a comprehensive, up-to-date summary of information on plant genome architecture and evolution in this review.

Key words: Genome size; Polyploidy; Whole-genome duplication; Transposable element; GC content.

Introduction

The origin and early evolution of land plants date back to the mid-Paleozoic, between approximately 480 and 360 million years ago (MYA) [1]. The ~ 350,000 extant land species exhibit remarkable diversity, including liverworts, hornworts, mosses, and vascular plants, with a wide range of complex organs and tissue systems. They are thought to have far-reaching consequences for the evolution of terrestrial organisms and the global environments [2]. The study of plant evolution has been conducted across several domains: the fossil record [3], morphology [4], and molecular approaches [5]. The development of genomic technologies has advanced the study of plant evolution, enhancing not only classification but also structural and functional exploration.

The first completed genome of *Arabidopsis thaliana* was published in 2000, with a total length of 115.4 Mb [6]. It is a small diploid genome and is regarded as a model system for structure and function studies of the plant genome due to its relatively simple genome content and growth cycle [6]. Whole-genome sequencing has made the complexity of the plant genome much clearer and more comprehensive. Through the comparative analyses of an increasing number of plant genome sequences and functional genomic datasets, we now have unprecedented insights into the impact of various factors on genome evolution. The purpose of this review is to provide an overview of our current understanding of plant genome size (GS) and composition. Based on our integrated data analysis and literature mining results, we focus on the variation in GS and composition among plants and the prevalence of polyploidy, and finally outline the impact of transposable elements (TEs) in plants.

Genome size

The total amount of DNA in the haploid or gametic nucleus of a species is referred to as the C-value or GS [7]. The GS of a species can be estimated using flow cytometry (FCM) [8] or by bioinformatic methods based on *k*-mer reads [9]. Among eukaryotes, GS varies approximately 64,000-fold [10], ranging from the smallest genome of the parasitic microsporidian *Encephalitozoon intestinalis* with a C-value of only 0.002 pg (2.3 Mb) [11] to the largest reliable GS estimated for the angiosperm *Paris japonica*, with a C-value of 152.23 pg (148,880 Mb) [12]. In plants, the Plant DNA C-DNA Database (<https://cval.ues.science.kew.org/>) [13] indicates that GS can vary 2440-fold, similar to the 3300-fold variation observed in animals [14]. Among plants, angiosperms are particularly notable for their variability, with GS varying more than 2000-fold. Species with large genomes are largely restricted to monocots, especially in the families Alliaceae, Asparagaceae, Liliaceae, Melanthiaceae, and Orchidaceae [15]. Although the GSs of gymnosperms are extremely prominent among seed plants, their variation is more conservative, with GS ranging from 2.25 pg (*Gnetum ula*) to 36 pg (*Pinus ayacahuite*), reflecting an approximately 16-fold difference (Figure 1) [13]. Compared with seed plants, seed-free plants usually display smaller GS distributions [16]. Ferns and lycophytes, among seed-free plants, exhibit overall larger and more variable GSs than bryophytes. In contrast to ferns and lycophytes, there appears to be a correlation between GS and chromosome number in mosses and possibly in all bryophytes [16–18]. Furthermore, due to the extinction of many important ancestral species and the fact that a large number of plant genomes remain unquantified, further exploration of plant GS distribution patterns requires more analysis.

Received: 18 May 2023; Revised: 23 August 2024; Accepted: 24 October 2024.

© The Author(s) 2024. Published by Oxford University Press and Science Press on behalf of the Beijing Institute of Genomics, Chinese Academy of Sciences / China National Center for Bioinformation and Genetics Society of China.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

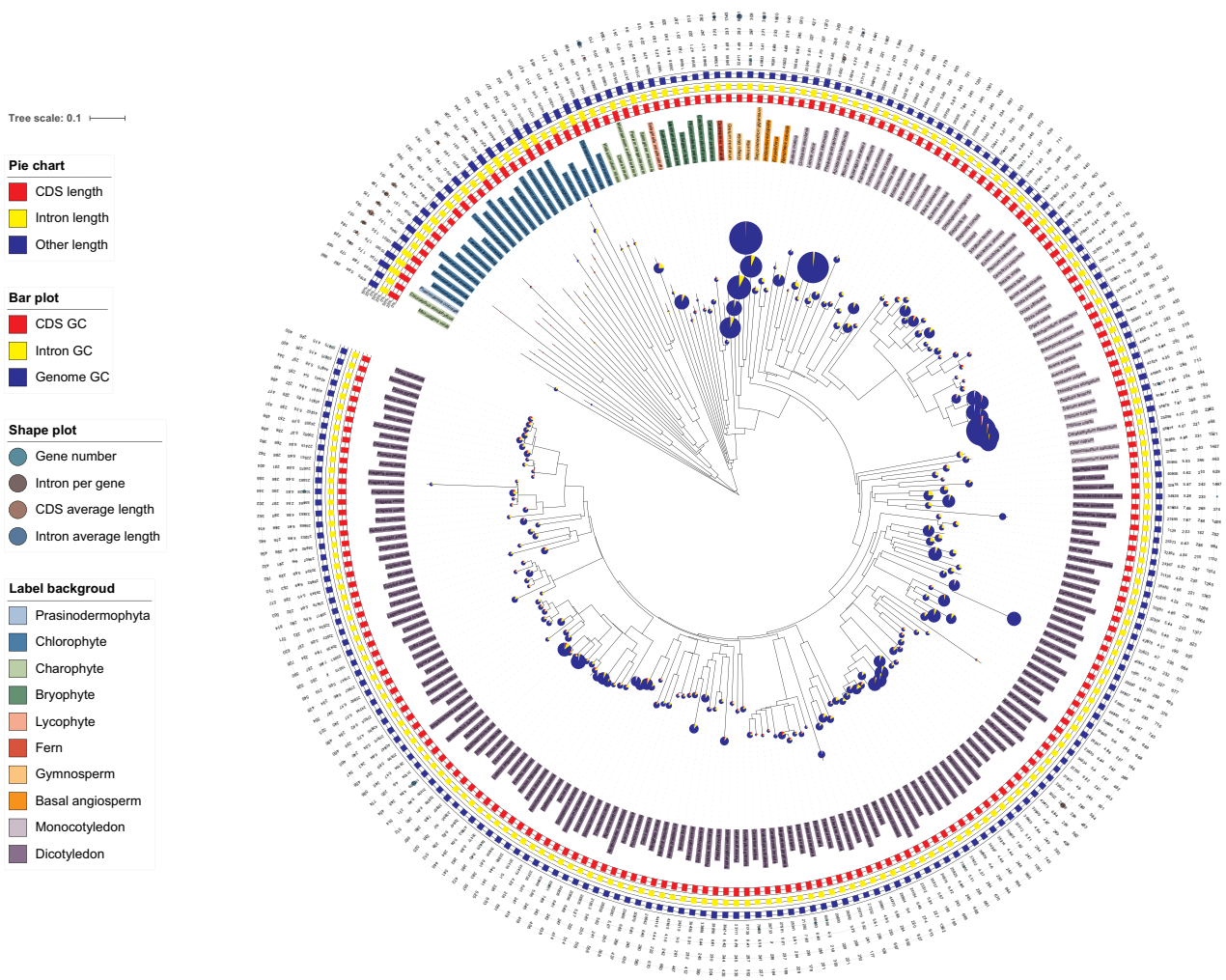


Figure 1 Global landscape of GSs across 234 representative plant species

The pie chart illustrates GSs consisting of CDSs, introns, and others. The bar plot shows GC contents of the CDSs, introns, and genomes. The shape plot shows the number of genes, average intron number per gene, average CDS length, and average intron length. Different classifications are represented with different background colors in the "label region". GS, genome size.

Consistent with the effects of polyploidization events, the expansion of plant GS enhances adaptation to specific environments [19]. Plant species with larger genomes typically have larger, sparser stomata, which limit stomatal conductance to CO_2 and water [20] and slow the stomatal response to environmental changes [21]. The primary mechanisms promoting higher GS in plants are mainly polyploidization events, including whole-genome duplication (WGD), and the accumulation of substantial numbers of repetitive elements [16]. However, it is clear that the trend in GS is unlikely to be unidirectional, the so-called "one-way ticket to genome obesity" [22], but rather bidirectional, experiencing both contractions and expansions. The most typical example is *Arabidopsis*, which has undergone multiple WGD events [23]. If GS changes followed a unidirectional trend, *Arabidopsis* would currently be 48-ploid instead of diploid [24]. This conclusion is also supported by GS results in gymnosperms, where the range of GS fold change variation remains one of the smallest among plants, despite the prevalence of bulky genomes and the fact that the GS expansion mechanism differs from that of angiosperms in that TE insertion events rather than polyploidization events are the main cause of their amplification [25].

There are several explanations for the phenomenon of plant genome downsizing, such as the tendency to produce smaller genomes given the nitrogen (N) and phosphorus (P) loads of nucleic acids [26,27] or the selection for smaller genomes to improve photosynthetic and water use efficiency [19,28]. In terms of molecular mechanisms, GS contraction in angiosperms may be closely related to uneven and illegitimate recombination, as well as double-strand break (DSB) repair involving DNA recombination [16,29]. From the results of existing studies, it seems that there is a dynamic mechanism to regulate GS in plants. Under environmental stress or survival pressure, plants prefer to increase GS through polyploidization and TE insertion to improve their environmental adaptability. When the fitness landscape is stable, plants seem to exhibit genome downsizing to reduce the energy cost of large genomes [30]. It is noteworthy that, as one of the few relict species that survived the mass extinction, the generally large GS and narrow GS range of gymnosperms has raised questions about the ultimate destination of GS evolution in these species. It is interesting to consider whether these species have fully adapted to their current environment and are experiencing genome contraction, or if they still intend to pursue genome expansion.

GC content variation

Typically, features such as gene size and GC content of the coding sequences (CDSs) have been shown to vary significantly both between different genomes and between different regions within the genome [31,32]. However, systematic studies in related areas have been scarce due to insufficient information on plant genomes. To explore this issue further, we obtained the genomes of 234 representative plant species with chromosome-level genome assemblies or unique evolutionary nodes from public databases. These plant species include 137 dicotyledons (22 in Fabales, 19 in Rosales, 12 in Solanales, 10 in Brassicales, nine in Malpighiales, and others), 51 monocotyledons (30 in Poales, five in Asparagales, four in Alismatales, three in Arecales, three in Dioscoreales, and others), three basal angiosperms, four gymnosperms, one fern, one lycophyte, six bryophytes, seven charophytes, 23 chlorophytes, and one prasinodermophyte (Table S1).

According to our results, the total gene size ranged from 9.88 Mb (74.84% of the GS) in chlorophytes to 710.04 Mb (7.43% of the GS) in gymnosperms. A high correlation coefficient was observed between gene size, intron size, and CDS size (gene size *vs.* intron size: $R = 0.9565$, $P = 4.2E-126$; gene size *vs.* CDS size: $R = 0.5818$, $P = 1.37E-22$). In many species of charophytes, gymnosperms, basal angiosperms, monocots, and dicots, the total intron size is more than half of the total gene size. Exon size appears to be limited by selective pressure (50 bp to 10 kb), whereas intron size varies widely within and between organisms (50 bp to 1 Mb). Most exon sizes in most species range from 100 bp to 1 kb, except for some species with broader exon sizes in chlorophytes and charophytes (*Chara braunii*). Intron size distributions tend to be unimodal (sharp or blunt) or bimodal (weak or a strong second peak) in plant species, and the size distribution patterns are highly variable among species. Further inspection reveals that intron size distributions are somewhat correlated with phylogeny, and species with closer evolutionary relationships have similar intron size distribution patterns. In general, according to our results, intron size distributions are relatively stable in dicots, although they are much more variable in other species, which is also consistent with previous conclusions [17,25].

From algae to higher plants, there is a trend of decreasing and then increasing GC content in plant genomes (Figure 2). *Raphidocelis subcapitata* in Chlorophyta has the highest genome GC content (71.62%), while *Cicer arietinum* in Dicotyledon has the lowest genome GC content (31.03%). Among them, plants in Chlorophyta have presented the highest diversity of genome GC content (from 40.38% to 71.62%, coefficient of variation = 0.1345). Compared with basal angiosperms, monocots and dicots have a higher diversity of genome GC content, and these observations identify the increase and decrease in the genome GC content of monocots and dicots. In addition, monocots have a higher average genome GC content than dicots (41.94% *vs.* 35.99%). Similar to the dynamics observed in the genome, the diversity of the GC content in genes, CDSs, and introns is similar. Most species have an increase in CDS GC content and a decrease in intron GC content compared with genome GC content, except for *Prasinoderma coloniale* in Prasinodermophyta and the genus *Ostreococcus* in Chlorophyta, which have higher intron GC content and lower CDS GC content. An

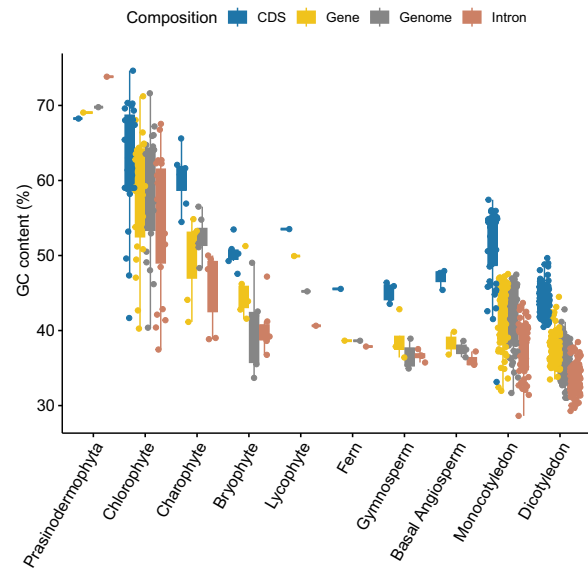


Figure 2 GC content results for different plant genome regions

A total of 234 representative plant species were included in the GC content comparison, comprising 137 dicotyledons, 51 monocotyledons, three basal angiosperms, four gymnosperms, one fern, one lycophyte, six bryophytes, seven charophytes, 23 chlorophytes, and one prasinodermophyte.

increased GC content distribution for exon and intron is observed in Poaceae species, which is consistent with previous results [33,34].

Several hypotheses have been proposed to explain the origin and evolution of changes in GC content, including selection for GC content, mutational bias, and GC-biased gene conversion (gBGC) [31,34–36]. gBGC is a recombination-related process that favors G/C over A/T bases during mismatch repair, and based on evidence for GC-biased enrichment of highly recombinant regions as well as GC-biased mismatch repair, it is increasingly argued that gBGC should be a major driver of changes in GC content [37–39]. Furthermore, the GC and GC3 contents of CDSs in dicotyledons show a unimodal distribution, whereas in monocot genomes, especially in grasses, they show a bimodal distribution. Interestingly, based on the results of our analysis, it seems that this phenomenon is not only present in monocots, but also in the green alga *Ulva mutabilis* with a bimodal distribution (Figure 3). It is worth noting that DNA sequence, as the ultimate carrier of genetic information, can be understood as a mapping of high-dimensional genetic information to lower dimensions. Therefore, highly compressed genomic features such as GC content may be the result of a combination of factors, and it is worth considering whether there is a highly robust mechanism involved.

Plant polyploidy

Diploids reproduce primarily by producing haploid gametes, which combine to form diploid offspring. In rare cases, polyploidy can be produced either by somatic duplication of chromosomes in fertilized eggs or by the production of unreduced gametes, the latter being the primary mechanism of polyploidy production [40]. Polyploidy is more prevalent in plants than in animals, although its distribution among plants

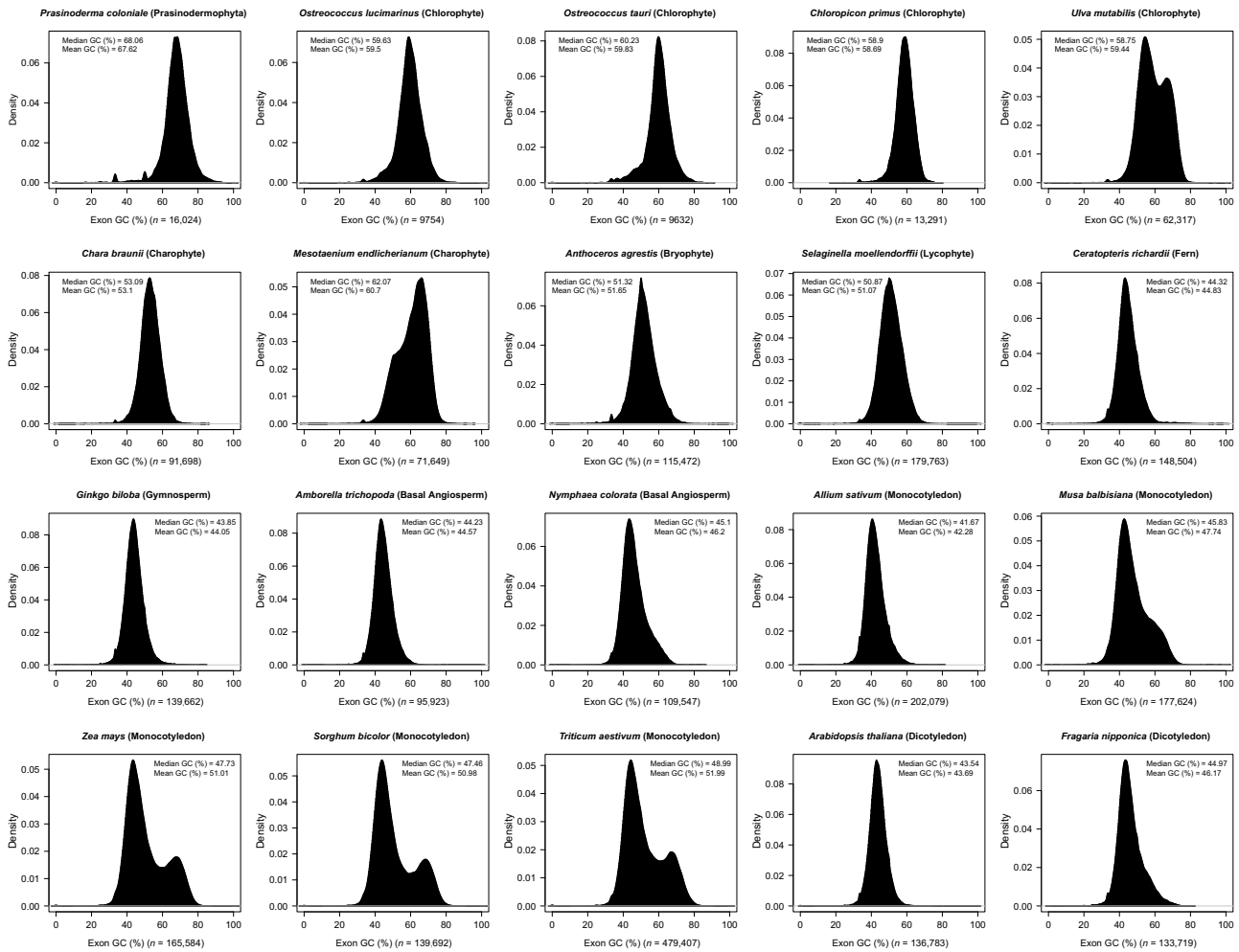


Figure 3 Distribution of GC content in the CDS region of representative plants
CDS, coding sequence.

is not uniform [41]. For example, the frequency of polyploidy in ferns is as high as 95% [42], while in angiosperms the frequency ranges from 30% to 70%, with dicots generally exhibiting a higher degree of polyploidy than monocots [43]. On the other hand, natural polyploids are particularly rare in gymnosperms and are absent in cycads (Cycadophyta) or ginkgos (Ginkgophyta), although polyploid plants of ginkgos can be artificially cultivated under certain conditions [44,45]. The prevalence of polyploids in plants is intriguing because polyploids are anticipated to be rare and evolutionarily short-lived due to the frequency-dependent reproductive disadvantages associated with being the minority cytotype when they first arise in a population [46].

Although there are multiple causes of plant polyploid formation in terms of cytological pathways (Figure 4), most polyploid cultivars have been domesticated by wild-to-crop and interploidy introgression, and these two typical events are also important for the formation of allopolyploid and autopolyploid, respectively [47,48]. For example, hexaploid wheat (*Triticum aestivum*, AABBDD) is derived from a hybridization between cultivated tetraploid wheat (*T. turgidum*, AABB) and the goat grass *Aegilops tauschii* (DD) [49]. The agriculturally important genus *Brassica* contains three diploid species, *Brassica rapa* (AA, $n = 10$) [50], *B. nigra* (BB, $n = 8$) [51], *B. oleracea* (CC, $n = 9$) [52], and three allopolyploid

species, *B. juncea* (AABB, $n = 18$) [53], *B. napus* (AACC, $n = 19$) [54], and *B. carinata* (BBCC, $n = 17$) [55]. Although previous studies have generally concluded that the number of autopolyploids is much lower than that of allopolyploids [56], ongoing studies suggest that the frequency and role of autopolyploidy in plants may be much higher [57,58].

Despite the large number of genes generated by polyploidy events that tend to be lost or non-functional, a growing body of evidence has shown that polyploidy plays a crucial role in enabling plants, especially angiosperms, in adapting to their environment and responding to stress, which may be an important driver of speciation [59–61]. In addition, polyploid species tend to show enhanced adaptations to salt stress [62], cold stress [63], or drought [64], and there appears to be a positive correlation between their frequency of distribution and latitude [65]. Although plants do not typically undergo polyploidization in stable environments, an excessively large genome can impose an additional energy burden on the organism. In times of high survival pressure, polyploidy events can lead to altered gene expression, epigenetic remodeling, or gene sub/neofunctionalization, providing a better buffer for plants [61]. It is worth noting that extant gymnosperms have also experienced several dramatic fluctuations in the Earth's habitats over a long period of time, raising the question of whether the extreme rarity of polyploidy in gymnosperms

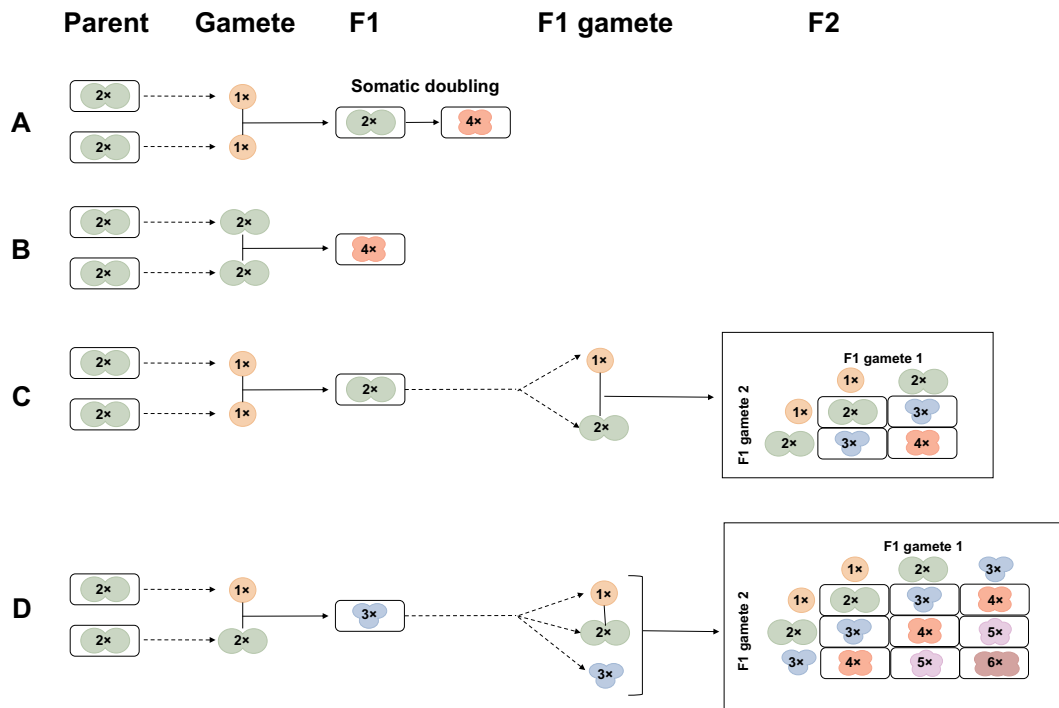


Figure 4 Different mechanisms of polyploid formation using tetraploid as an example

A. Somatic cell doubling. **B.** Fusion of unreduced gametes. **C.** A certain percentage of aneuploid gametes may be produced in the F1 offspring. **D.** The triploid bridge scenario in which the intermediate triploid produces unreduced gametes, leading to polyploidy.

implies that these species possess a fundamentally different mechanism for environmental adaptation compared with angiosperms.

The genome assembly of polyploids, especially homopolyploids, is challenging, because traditional diploid genome assembly strategies tend to ignore differences between haplotypes, resulting in chimeric haploid genome assembly, which can lead to significant errors for polyploids [66,67]. There are three main ideas to solve this problem: physical segregation methods, *de novo* haplotype assembly, and alignment-based phasing [68]. Although the physical segregation strategy is optimal for obtaining haplotypic gametophytes for sequencing, it has a significant human cost and presents several technical drawbacks [69]. The alignment-based polyploid phasing strategy, on the other hand, is based on well-assembled reference sequences and relies on sequence comparison, recording of phase-informative positions, optimization of objective functions [*e.g.*, minimum error correction (MEC)], and graph partitioning and clustering to construct haplotype genomes [70,71]. Alignment-based phasing has the advantage of being less computationally intensive, but the results obtained are highly dependent on the quality and structural complexity of the reference genome [72]. The strategy of *de novo* assembly is an attempt to reconstruct different haplotypes and resolve the structure of the genome using chromosomal long-range interactions (Hi-C) and/or parental information in the absence of reference sequences, and string graphs are also a common structure of existing algorithms under this strategy [73–76].

WGD

WGD is a ubiquitous feature of plant genomes, contributing to variation in both GS and gene content. Although WGD has

occurred in all major clades of terrestrial plants, its frequency is not fixed. More recent events have been observed in angiosperms, which are associated with increased rates of angiosperm diversification after the Cretaceous [77]. The two major clades of angiosperms, eudicots and monocots, both experienced paleopolyploidization events early in their evolutionary history, termed gamma (γ) and tau (τ), dating back to ~ 120 MYA and ~ 66 MYA, respectively [78,79]. However, the occurrence of recent WGD events in eudicots and monocots was completely independent. Two WGD events in monocots (ρ and σ) were inferred to have preceded the diversification of cereals and grasses. And studies have revealed two recent WGDs (α and β) supported by *Arabidopsis* within the *Crucifer* lineage, as well as a triplication event (γ) likely shared by all core eudicots [78,80,81].

The ancestors of angiosperm lineages underwent a significant number of WGD events, as inferred from studies of gene duplication mutation patterns (Ks analysis), multi-tAxon paleopolyploidy search (MAPS analysis), and a combination of maximum likelihood and Bayesian approaches [82–84]. It should be noted that due to algorithmic or modeling differences, there may be variations in the estimation of differentiation time and speculation of WGD events among these methods. On the other hand, in gymnosperms, although studies based on transcriptomic data suggest that there was a common WGD event that occurred during the divergence of seed plants [85], this conclusion is controversial for angiosperms, and it is also intriguing whether there are lineage-specific WGD events in gymnosperms [86]. Recently, the nearly complete genome assemblies of *Ginkgo*, *Cycas*, and *Taxus* provided the syntenic genes to confirm a common seed plant WGD and no lineage-specific WGDs, putting these controversies to rest for the time being [25,87,88]. Compared with other land plants, the genomic characteristics of seed-free plants are significantly different, *e.g.*,

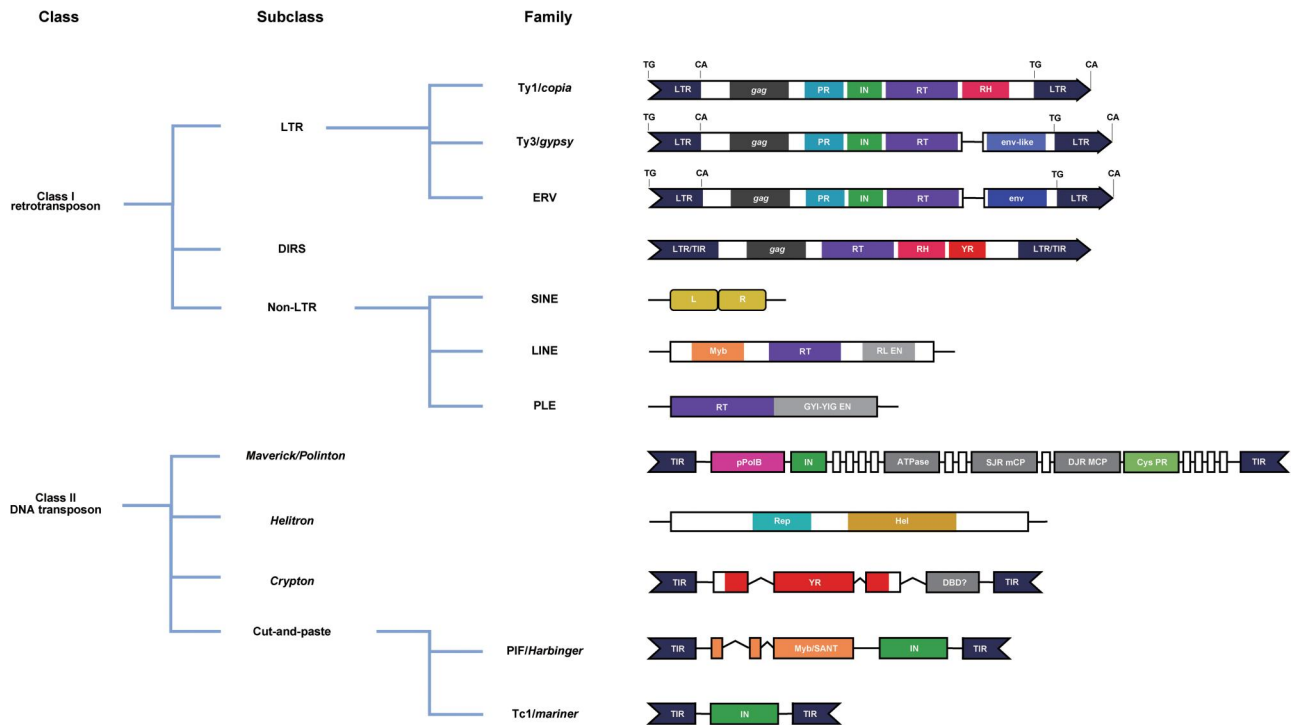


Figure 5 Classification and structural characteristics of major TEs in plants

TE, transposable element; LTR, long terminal repeat; DIRS, *Dictyostelium* intermediate repetitive sequence; ERV, endogenous retrovirus; SINE, short interspaced nuclear element; LINE, long interspaced nuclear element; PLE, *Penelope*-like element; PIF, *P* instability factor; PR, protease; IN, integrase; RT, reverse transcriptase; RH, restriction endonuclease; YR, tyrosine recombinase; TIR, terminal inverted repeats; L, left; R, right; RL EN, RAG-like (recombination activating gene) enzyme; GYI-YIG EN, GYI-YIG endonuclease; SJR mCP, single jelly-roll minor capsid protein; DJR MCP, double jelly-roll major capsid protein; Cys PR, cysteine protease; Rep, replication protein; Hel, helicase; DBD, DNA binding domain.

hornworts, *Selaginella*, and most liverworts do not have WGD events and appear to have lower methylation levels [17]. In contrast, there has been at least one well-supported WGD in the history of most mosses, ferns, and homosporous lycophytes [89]. It is important to note that some key clades currently lack reference genomes, so further relevant data are needed for interpretation.

The main consequences behind WGD and polyploidy function involve the activation of TEs and epigenetic changes that subsequently lead to an increase in mutation rates and regulatory changes in genes and their genomic structure, ultimately leading to the diversification or the emergence of new phenotypes [77,90,91]. WGD is one of the major sources of genetic innovation in land plants, especially in angiosperms [90,92]. A commonly observed effect of WGD is the “gigas effect”, which refers to an increase in the size and robustness of plant traits [93]. The gigas effect is thought to be the result of polyploids having greater amounts of DNA, resulting in larger cells and translating into larger tissues and organs. There is evidence indicating that phenotypes may undergo deterioration or alteration in subsequent generations following WGD [94]. However, it is important to note that not all WGD events resulted in diversification. In particular, 46 out of 106 WGD events did not exhibit significant interactions with species formation in angiosperms [84].

TEs in plants

TEs are mobile DNA sequences that can replicate independently within the host genome. Typically, they range in length from 100 bp to 10,000 bp [95]. Our understanding of the

structure, organization, and translocation mechanisms of TEs has advanced considerably since their initial discovery in maize by Barbara McClintock [96]. Based on how these elements move through genomes, TEs can be divided into two main categories: retrotransposons (class I elements) and DNA transposons (class II elements). Retrotransposons use an encoded reverse transcriptase to synthesize DNA from RNA, increasing the copy number of the transposon. This mode of transposon movement is commonly referred to as “copy and paste”. DNA transposons utilize a “cut-and-paste” mechanism to jump from one locus to another within a genome (Figure 5) [97].

Class I retrotransposons can be further categorized into long terminal repeat-retrotransposons (LTR-RTs), non-LTR retrotransposons, and *Dictyostelium* intermediate repetitive sequences (DIRSs), with LTR-RTs being predominant in most plant species [98]. For example, LTR-RTs can comprise up to 80% of the maize genome [99]. LTR-RTs resemble retroviruses both in structure and mechanism and are flanked by conserved palindromic sequences (5'-TG ... CA-3') [100], ranging from a few hundred base pairs to more than 5 kb [101]. Most LTR-RTs contain two open reading frames (ORFs), *gag* and *pol*, where the *gag* ORF encodes the structural protein of the virus-like particle, while the *pol* ORF encodes aspartic proteinase (AP), reverse transcriptase (RT), RNase H (RH), and integrase (IN) [102]. New insertion of LTR-RTs creates a short repeat at the ends of the element, called a target site duplication (TSD), which is 4–6 bp long [103]. In addition, some plant LTR-RTs may contain additional open reading frames (aORFs), although their function is not yet clear [104]. Non-LTR retrotransposons are less

abundant in plants, and can be divided into three major superfamilies: long interspaced nuclear elements (LINEs), short interspaced nuclear elements (SINEs), and *Penelope*-like elements (PLEs). Non-LTR retrotransposons lack LTRs and are transcribed from an internal promoter, and these elements can replicate themselves in the absence of the INT domain. The RT structural domain initiates DNA synthesis from the poly-A tail of the non-LTR retrotransposon transcript and ultimately ligates the newly synthesized DNA end to the insertion site [105].

Like class I transposons, class II transposons (DNA transposons) can be further subdivided into four subclasses: cut-and-paste, *Helitron*, *Crypton*, and *Maverick/Polinton* [106]. Cut-and-paste type DNA transposons, also known as terminal inverted repeat (TIR) TEs, tend to be the most abundant DNA transposons in plants and are characterized by autonomous copy-encoded transposases and terminally inverted repeats [107]. It is worth noting that these non-autonomous elements can retain binding sites recognized by autonomous transposons during transposition, resulting in an extensive MITE family that tends to have a greater ability to proliferate [108]. The most important feature of *Helitrons* is that they transpose with a “rolling cycle” mechanism via a transposase. This means that instead of a TIR, a strand of the original element is copied directly to the target site, a mechanism that is distinct from other DNA transposons [109]. *Cryptons* are relatively rare in eukaryotes and are also structurally similar to bacterial and archaeal insertion sequences [110]. *Maverick/Polintons* are widespread in eukaryotes. Although they are rare in number, they tend to be relatively large, often 15 kb or more in length. *Maverick/Polintons* are replicated by direct synthesis of DNA copies and are therefore known as self-synthesizing DNA transposons [95].

As with polyploidy events, the activity of TEs has played a major role in the evolution of plant genomes. Very few eukaryotic species are completely devoid of TEs [11], and most organisms that experience waves of TE mobilization also appear to contribute to the resizing and shaping of their genomes [111]. The average fraction of the genome occupied by TEs is ~ 50% of the total genome, with a range from 10% to more than 85% [112]. In general, there is a negative correlation between GS and TE diversity in terrestrial plants. Typically, angiosperms tend to have smaller genomes than gymnosperms, but the TE landscape is more diverse, whereas most gymnosperms are dominated by LTR-RTs [95].

TEs have long been regarded as “junk DNA” and are typically masked during genome annotation to avoid influencing the annotation results. However, as research has advanced, it has become increasingly evident that TEs play an increasingly important role in both the structural and functional domains [113]. For example, TEs can enhance gene expression and produce phenotypic differences when integrated into the *cis*-regulatory region [114–116]. Additionally, they can function as “control elements” to mediate large-scale chromosomal rearrangements [117,118], together with extensive horizontal transfer [119]. Although TEs maintain an extremely high proportion in some plant genomes, their massive proliferation does not seem to be unconditionally tolerated by plants. As mentioned above, the enlarged genomes resulting from overactive TEs can increase the energetic burden on the organism. Available data suggest that DNA methylation mediated by the RNA-directed DNA methylation (RdDM) pathway and histone modification appear to be important

mechanisms for regulating the expression of TEs [120–123]. Research on plant TEs remains in its early stages, encompassing the classification and identification of novel TE types as well as understanding their evolving roles in genome evolution and functional regulation, all of which require further investigation.

Conclusion and future perspectives

As sequencing data continue to be published, plants have already shown great variation in GS, composition, and other important characteristics. Polyploidy and its associated WGD, along with the abundance of TEs, are the main reasons for the differences in their GSs. Based on the available results, plant adaptation and interactions with their living environments should be considered as significant drivers of plant genome evolution; this evolutionary process does not seem to be consistently stable and gradual. The significant differences in the genomic characteristics among various plants may indicate distinct strategies adopted by different plants to adapt to their respective environments. Conversely, these differences in the selection of environmental adaptation strategies may ultimately lead to divergent genomic landscapes. A typical example is the varying preferences for polyploidy and TEs between gymnosperms and angiosperms. To take another specific example, the average length of introns in gymnosperms is significantly greater than that in angiosperms. In cycads, for instance, the proportion of introns exceeds 50% of the genome, whereas in *Arabidopsis*, more than 90% of intron lengths are less than 1 kb [87]. Thus, whether in terms of energy loss through transcription and splicing or transcriptional regulation at three-dimensional or higher levels of gene organization, it is almost certain that genomic landscapes with such disparate characteristics will also have very different molecular regulatory mechanisms. In subsequent studies, it is essential to gather more genomic data from representative species, especially lower plants or key clades, to construct a complete evolutionary atlas of plant genomes. Additionally, the study of genome annotation and regulatory mechanisms in plants may require more attention.

CRedit author statement

Bing He: Writing – original draft, Writing – review & editing. **Wanfei Liu:** Writing – original draft, Software, Formal analysis. **Jianyang Li:** Software, Formal analysis. **Siwei Xiong:** Software, Formal analysis. **Jing Jia:** Software, Formal analysis. **Qiang Lin:** Writing – review & editing, Investigation. **Hailin Liu:** Writing – original draft, Writing – review & editing, Validation. **Peng Cui:** Conceptualization, Writing – review & editing, Supervision. All authors have read and approved the final manuscript.

Supplementary material

Supplementary material is available at *Genomics, Proteomics & Bioinformatics* online (<https://doi.org/10.1093/gpbjnl/qzae078>).

Competing interests

The authors declare no competing interests.

Acknowledgments

This work was supported by special funds for the science technology innovation and industrial development of Shenzhen Dapeng New District, China (Grant No. PT202101-02), the National Natural Science Foundation of China (Grant No. 32072101), the National Key Projects in Agricultural Biological Breeding, China (Grant No. 2022ZD0401702), and the Guangdong Basic and Applied Basic Research Foundation, China (Grant No. 2019A1515111150).

ORCID

0000-0001-5859-5754 (Bing He)
 0000-0002-6637-6807 (Wanfei Liu)
 0009-0002-4884-466X (Jianyang Li)
 0009-0004-6763-2974 (Siwei Xiong)
 0000-0003-2700-088X (Jing Jia)
 0000-0003-2882-7758 (Qiang Lin)
 0000-0002-4180-3673 (Hailin Liu)
 000-0003-2989-571X (Peng Cui)

References

- [1] Kenrick P, Crane PR. The origin and early evolution of plants on land. *Nature* 1997;389:33–9.
- [2] Feijen FAA, Vos RA, Nuytinck J, Merckx VSFT. Evolutionary dynamics of mycorrhizal symbiosis in land plant diversification. *Sci Rep* 2018;8:10698.
- [3] Jackson JBC, Erwin DH. What can we learn about ecology and evolution from the fossil record? *Trends Ecol Evol* 2006;21:322–8.
- [4] Jaramillo MA, Kramer EM. The role of developmental genetics in understanding homology and morphological evolution in plants. *Int J Plant Sci* 2007;168:61–72.
- [5] Schreiber M, Stein N, Mascher M. Genomic approaches for studying crop evolution. *Genome Biol* 2018;19:104.
- [6] Kaul S, Koo HL, Jenkins J, Rizzo M, Rooney T, Tallon LJ, et al. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 2000;408:796–815.
- [7] Dolezel J, Bartos J, Voglmayr H, Greilhuber J. Nuclear DNA content and genome size of trout and human. *Cytometry A* 2003;51:127–8.
- [8] Dolezel J, Bartos J. Plant DNA flow cytometry and estimation of nuclear genome size. *Ann Bot* 2005;95:99–110.
- [9] Sun H, Ding J, Piednoel M, Schneeberger K. *findGSE*: estimating genome size variation within human and *Arabidopsis* using *k*-mer frequencies. *Bioinformatics* 2018;34:550–7.
- [10] Pellicer J, Fay MF, Leitch IJ. The largest eukaryotic genome of them all? *Bot J Linn Soc* 2010;164:10–5.
- [11] Corradi N, Pombert JF, Farinelli L, Didier ES, Keeling PJ. The complete sequence of the smallest known nuclear genome from the microsporidian *Encephalitozoon intestinalis*. *Nat Commun* 2010;1:77.
- [12] Ayala FJ, Fitch WM, Clegg MT. Variation and evolution in plants and microorganisms: toward a new synthesis 50 years after Stebbins. *Proc Natl Acad Sci U S A* 2000;97:6941–4.
- [13] Pellicer J, Leitch IJ. The Plant DNA C-values database (release 7.1): an updated online repository of plant genome size data for comparative studies. *New Phytol* 2020;226:301–5.
- [14] Gregory TR, Nicol JA, Tamm H, Kullman B, Kullman K, Leitch IJ, et al. Eukaryotic genome size databases. *Nucleic Acids Res* 2007;35:D332–8.
- [15] Leitch IJ, Kahandawala I, Suda J, Hanson L, Ingrouille MJ, Chase MW, et al. Genome size diversity in orchids: consequences and evolution. *Ann Bot* 2009;104:469–81.
- [16] Pellicer J, Hidalgo O, Dodsworth S, Leitch IJ. Genome size diversity and its impact on the evolution of land plants. *Genes (Basel)* 2018;9:88.
- [17] Szovenyi P, Gunadi A, Li FW. Charting the genomic landscape of seed-free plants. *Nat Plants* 2021;7:554–65.
- [18] Clark J, Hidalgo O, Pellicer J, Liu HM, Marquardt J, Robert Y, et al. Genome evolution of ferns: evidence for relative stasis of genome size across the fern phylogeny. *New Phytol* 2016;210:1072–82.
- [19] Wang X, Morton JA, Pellicer J, Leitch IJ, Leitch AR. Genome downsizing after polyploidy: mechanisms, rates and selection pressures. *Plant J* 2021;107:1003–15.
- [20] Franks PJ, Farquhar GD. The effect of exogenous abscisic acid on stomatal development, stomatal mechanics, and leaf gas exchange in *Tradescantia virginiana*. *Plant Physiol* 2001;125:935–42.
- [21] Roddy AB, Theroux-Rancourt G, Abbo T, Benedetti JW, Brodersen CR, Castro M, et al. The scaling of genome size and cell size limits maximum rates of photosynthesis with implications for ecological strategies. *Int J Plant Sci* 2020;181:75–87.
- [22] Bennetzen JL, Kellogg EA. Do plants have a one-way ticket to genomic obesity? *Plant Cell* 1997;9:1509–14.
- [23] Van de Peer Y, Mizrachi E, Marchal K. The evolutionary significance of polyploidy. *Nat Rev Genet* 2017;18:411–24.
- [24] Wendel JF. The wondrous cycles of polyploidy in plants. *Am J Bot* 2015;102:1753–6.
- [25] Liu H, Wang X, Wang G, Cui P, Wu S, Ai C, et al. The nearly complete genome of *Ginkgo biloba* illuminates gymnosperm evolution. *Nat Plants* 2021;7:748–56.
- [26] Hessen DO, Jeyasingh PD, Neiman M, Weider LJ. Genome streamlining and the elemental costs of growth. *Trends Ecol Evol* 2010;25:75–80.
- [27] Guignard MS, Nichols RA, Knell RJ, Macdonald A, Romila CA, Trimmer M, et al. Genome size and ploidy influence angiosperm species' biomass under nitrogen and phosphorus limitation. *New Phytol* 2016;210:1195–206.
- [28] Lawson T, Violet-Chabrand S. Speedy stomata, photosynthesis and plant water use efficiency. *New Phytol* 2019;221:93–8.
- [29] Schubert I, Vu GTH. Genome stability and evolution: attempting a holistic view. *Trends Plant Sci* 2016;21:749–57.
- [30] Fragata I, Blanckaert A, Dias Louro MA, Liberles DA, Bank C. Evolution in the light of fitness landscape theory. *Trends Ecol Evol* 2019;34:69–82.
- [31] Singh R, Ming R, Yu Q. Comparative analysis of GC content variations in plant genomes. *Trop Plant Biol* 2016;9:136–49.
- [32] He B, Dong H, Jiang C, Cao F, Tao S, Xu LA. Analysis of codon usage patterns in *Ginkgo biloba* reveals codon usage tendency from A/U-ending to G/C-ending. *Sci Rep* 2016;6:35927.
- [33] Tatarinova TV, Alexandrov NN, Bouck JB, Feldmann KA. GC3 biology in corn, rice, sorghum and other grasses. *BMC Genomics* 2010;11:308.
- [34] Glemin S, Clement Y, David J, Ressayre A. GC content evolution in coding regions of angiosperm genomes: a unifying hypothesis. *Trends Genet* 2014;30:263–70.
- [35] Sueoka N. Directional mutation pressure and neutral molecular evolution. *Proc Natl Acad Sci U S A* 1988;85:2653–7.
- [36] Duret L, Galtier N. Biased gene conversion and the evolution of mammalian genomic landscapes. *Annu Rev Genomics Hum Genet* 2009;10:285–311.
- [37] Webster MT, Smith NGC, Hultin-Rosenberg L, Arndt PF, Ellegren H. Male-driven biased gene conversion governs the evolution of base composition in human alu repeats. *Mol Biol Evol* 2005;22:1468–74.
- [38] Spencer CCA. Human polymorphism around recombination hotspots. *Biochem Soc T Trans* 2006;34:535–6.
- [39] Lassalle F, Perian S, Bataillon T, Nesme X, Duret L, Daubin V. GC-content evolution in bacterial genomes: the biased gene conversion hypothesis expands. *PLoS Genet* 2015;11:e1004941.

- [40] Moghe GD, Shiu SH. The causes and molecular consequences of polyploidy in flowering plants. *Ann N Y Acad Sci* 2014; 1320:16–34.
- [41] Wood TE, Takebayashi N, Barker MS, Mayrose I, Greenspoon PB, Rieseberg LH. The frequency of polyploid speciation in vascular plants. *Proc Natl Acad Sci U S A* 2009;106:13875–9.
- [42] Grant V. Plant speciation, 2nd ed. New York: Columbia University Press; 1981.
- [43] Masterson J. Stomatal size in fossil plants: evidence for polyploidy in majority of angiosperms. *Science* 1994;264:421–4.
- [44] Ahuja MR. Polyploidy in gymnosperms: revisited. *Silvae Genet* 2005;54:59–69.
- [45] Smarda P, Vesely P, Smerda J, Bures P, Knapk O, Chytra M. Polyploidy in a ‘living fossil’ *Ginkgo biloba*. *New Phytol* 2016; 212:11–4.
- [46] Porturas LD, Anneberg TJ, Cure AE, Wang S, Althoff DM, Segraves KA. A meta-analysis of whole genome duplication and the effects on flowering traits in plants. *Am J Bot* 2019;106:469–76.
- [47] Soltis DE, Albert VA, Leebens-Mack J, Bell CD, Paterson AH, Zheng CF, et al. Polyploidy and angiosperm diversification. *Am J Bot* 2009;96:336–48.
- [48] Ramsey J, Schemske DW. Neopolyploidy in flowering plants. *Annu Rev Ecol Syst* 2002;33:589–639.
- [49] Salamini F, Ozkan H, Brandolini A, Schafer-Pregl R, Martin W. Genetics and geography of wild cereal domestication in the Near East. *Nat Rev Genet* 2002;3:429–41.
- [50] Wang X, Wang H, Wang J, Sun R, Wu J, Liu S, et al. The genome of the mesopolyploid crop species *Brassica rapa*. *Nat Genet* 2011;43:1035–157.
- [51] Paritosh K, Yadava SK, Singh P, Bhayana L, Mukhopadhyay A, Gupta V, et al. A chromosome-scale assembly of allotetraploid *Brassica juncea* (AABB) elucidates comparative architecture of the A and B genomes. *Plant Biotechnol J* 2021; 19:602–14.
- [52] Liu S, Liu Y, Yang X, Tong C, Edwards D, Parkin IAP, et al. The *Brassica oleracea* genome reveals the asymmetrical evolution of polyploid genomes. *Nat Commun* 2014;5:3930.
- [53] Yang J, Liu D, Wang X, Ji C, Cheng F, Liu B, et al. The genome sequence of allopolyploid *Brassica juncea* and analysis of differential homoeolog gene expression influencing selection. *Nat Genet* 2016;48:1225–32.
- [54] Chalhoub B. Early allopolyploid evolution in the post-Neolithic *Brassica napus* oilseed genome. *Science* 2014; 345:950–3.
- [55] Song X, Wei Y, Xiao D, Gong K, Sun P, Ren Y, et al. *Brassica carinata* genome characterization clarifies U’s triangle model of evolution and polyploidy in *Brassica*. *Plant Physiol* 2021; 186:388–406.
- [56] Parisod C, Holderegger R, Brochmann C. Evolutionary consequences of autopolyploidy. *New Phytol* 2010;186:5–17.
- [57] Soltis DE, Soltis PS, Schemske DW, Hancock JF, Thompson JN, Husband BC, et al. Autopolyploidy in angiosperms: have we grossly underestimated the number of species? *Taxon* 2007;56:13–30.
- [58] Barker MS, Arrigo N, Baniaga AE, Li Z, Levin DA. On the relative abundance of autopolyploids and allopolyploids. *New Phytol* 2016;210:391–8.
- [59] Van de Peer Y, Maere S, Meyer A. The evolutionary significance of ancient genome duplications. *Nat Rev Genet* 2009; 10:725–32.
- [60] Segraves KA. The effects of genome duplications in a community context. *New Phytol* 2017;215:57–69.
- [61] Van de Peer Y, Ashman TL, Soltis PS, Soltis DE. Polyploidy: an evolutionary and ecological force in stressful times. *Plant Cell* 2021;33:11–26.
- [62] Zhu H, Zhao S, Lu X, He N, Gao L, Dou J, et al. Genome duplication improves the resistance of watermelon root to salt stress. *Plant Physiol Biochem* 2018;133:11–21.
- [63] Lourkisti R, Froelicher Y, Herbette S, Morillon R, Tomi F, Gibernau M, et al. Triploid citrus genotypes have a better tolerance to natural chilling conditions of photosynthetic capacities and specific leaf volatile organic compounds. *Front Plant Sci* 2020;11:330.
- [64] Wang Y, Chen G, Zeng F, Han Z, Qiu CW, Zeng M, et al. Molecular evidence for adaptive evolution of drought tolerance in wild cereals. *New Phytol* 2023;237:497–514.
- [65] Rice A, Smarda P, Novosolov M, Drori M, Glick L, Sabath N, et al. The global biogeography of polyploid plants. *Nat Ecol Evol* 2019;3:265–73.
- [66] Schrunner SD, Mari RS, Ebler J, Rautiainen M, Seillier L, Reimer JJ, et al. Haplotype threading: accurate polyploid phasing from long reads. *Genome Biol* 2020;21:252.
- [67] Tang HB. Disentangling a polyploid genome. *Nat Plants* 2017; 3:688–9.
- [68] Saada OA, Friedrich A, Schacherer J. Towards accurate, contiguous and complete alignment-based polyploid phasing algorithms. *Genomics* 2022;114:110369.
- [69] Shi D, Wu J, Tang H, Yin H, Wang H, Wang R, et al. Single-pollen-cell sequencing for gamete-based phased diploid genome assembly in plants. *Genome Res* 2019;29:1889–99.
- [70] Moeinzadeh MH, Yang J, Muzychenko E, Gallone G, Heller D, Reinert K, et al. Ranbow: a fast and accurate method for polyploid haplotype reconstruction. *PLoS Comput Biol* 2020; 16:e1007843.
- [71] Saada OA, Tsouris A, Eberlein C, Friedrich A, Schacherer J. nPhase: an accurate and contiguous phasing method for polyploids. *Genome Biol* 2021;22:126.
- [72] Guk JY, Jang MJ, Choi JW, Lee YM, Kim S. *De novo* phasing resolves haplotype sequences in complex plant genomes. *Plant Biotechnol J* 2022;20:1031–41.
- [73] Koren S, Rhie A, Walenz BP, Dilthey AT, Bickhart DM, Kingan SB, et al. *De novo* assembly of haplotype-resolved genomes with trio binning. *Nat Biotechnol* 2018;36:1174–82.
- [74] Cheng H, Concepcion GT, Feng X, Zhang H, Li H. Haplotype-resolved *de novo* assembly using phased assembly graphs with hifiasm. *Nat Methods* 2021;18:170–5.
- [75] Cheng H, Jarvis ED, Fedrigo O, Koepfli KP, Urban L, Gemmill NJ, et al. Haplotype-resolved assembly of diploid genomes without parental data. *Nat Biotechnol* 2022;40:1332–5.
- [76] Kronenberg ZN, Rhie A, Koren S, Concepcion GT, Peluso P, Munson KM, et al. Extended haplotype-phasing of long-read *de novo* genome assemblies using Hi-C. *Nat Commun* 2021;12:1935.
- [77] Soltis PS, Soltis DE. Ancient WGD events as drivers of key innovations in angiosperms. *Curr Opin Plant Biol* 2016;30:159–65.
- [78] Jaillon O, Aury JM, Noel B, Policriti A, Clepet C, Casagrande A, et al. The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature* 2007;449:463–5.
- [79] Tang H, Bowers JE, Wang X, Paterson AH. Angiosperm genome comparisons reveal early polyploidy in the monocot lineage. *Proc Natl Acad Sci U S A* 2010;107:472–7.
- [80] Lyons E, Pedersen B, Kane J, Alam M, Ming R, Tang H, et al. Finding and comparing syntenic regions among *Arabidopsis* and the outgroups papaya, poplar, and grape: CoGe with Rosids. *Plant Physiol* 2008;148:1772–81.
- [81] Bowers JE, Chapman BA, Rong JK, Paterson AH. Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. *Nature* 2003;422:433–8.
- [82] Jiao Y, Wickett NJ, Ayyampalayam S, Chanderbali AS, Landherr L, Ralph PE, et al. Ancestral polyploidy in seed plants and angiosperms. *Nature* 2011;473:97–100.
- [83] Li Z, Baniaga AE, Sessa EB, Scascitelli M, Graham SW, Rieseberg LH, et al. Early genome duplications in conifers and other seed plants. *Sci Adv* 2015;1:e1501084.
- [84] Landis JB, Soltis DE, Li Z, Marx HE, Barker MS, Tank DC, et al. Impact of whole-genome duplication events on diversification rates in angiosperms. *Am J Bot* 2018;105:348–63.
- [85] One Thousand Plant Transcriptomes Initiative. One thousand plant transcriptomes and the phylogenomics of green plants. *Nature* 2019;574:679–85.

- [86] Roodt D, Lohaus R, Sterck L, Swanepoel RL, Van de Peer Y, Mizrahi E. Evidence for an ancient whole genome duplication in the cycad lineage. *PLoS One* 2017;12:e0184454.
- [87] Liu Y, Wang S, Li L, Yang T, Dong S, Wei T, et al. The *Cycas* genome and the early evolution of seed plants. *Nat Plants* 2022; 8:389–401.
- [88] Xiong X, Gou J, Liao Q, Li Y, Zhou Q, Bi G, et al. The *Taxus* genome provides insights into paclitaxel biosynthesis. *Nat Plants* 2021;7:1026–36.
- [89] Li FW, Brouwer P, Carretero-Paulet L, Cheng S, de Vries J, Delaux PM, et al. Fern genomes elucidate land plant evolution and cyanobacterial symbioses. *Nat Plants* 2018;4:460–72.
- [90] Doyle JJ, Coate JE. Polyploidy, the nucleotype, and novelty: the impact of genome doubling on the biology of the cell. *Int J Plant Sci* 2019;180:1–52.
- [91] Vicent CM, Casacuberta JM. Impact of transposable elements on polyploid plant genomes. *Ann Bot* 2017;120:195–207.
- [92] Cheng SF, Xian WF, Fu Y, Marin B, Keller J, Wu T, et al. Genomes of subaerial Zygnematophyceae provide insights into land plant evolution. *Cell* 2019;179:1057–67.e4.
- [93] Clo J, Kolar F. Short- and long-term consequences of genome doubling: a meta-analysis. *Am J Bot* 2021;108:2315–22.
- [94] Oswald BP, Nuismer SL. Neopolyploidy and diversification in *Heuchera grossulariifolia*. *Evolution* 2011;65:1667–79.
- [95] Wells JN, Feschotte C. A field guide to eukaryotic transposable elements. *Annu Rev Genet* 2020;54:539–61.
- [96] Mc CB. The origin and behavior of mutable loci in maize. *Proc Natl Acad Sci U S A* 1950;36:344–55.
- [97] Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhoub B, et al. A unified classification system for eukaryotic transposable elements. *Nat Rev Genet* 2007;8:973–82.
- [98] Ramakrishnan M, Satish L, Sharma A, Vinod KK, Emamverdian A, Zhou MB, et al. Transposable elements in plants: recent advancements, tools and prospects. *Plant Mol Biol Rep* 2022;40:628–45.
- [99] Jiao Y, Peluso P, Shi J, Liang T, Stitzer MC, Wang B, et al. Improved maize reference genome with single-molecule technologies. *Nature* 2017;546:524–7.
- [100] Ou S, Jiang N. LTR_retriever: a highly accurate and sensitive program for identification of long terminal repeat retrotransposons. *Plant Physiol* 2018;176:1410–22.
- [101] Schietgat L, Vens C, Cerri R, Fischer CN, Costa E, Ramon J, et al. A machine learning based framework to identify and classify long terminal repeat retrotransposons. *PLoS Comput Biol* 2018; 14:e1006097.
- [102] Piednoel M, Carrete-Vega G, Renner SS. Characterization of the LTR retrotransposon repertoire of a plant clade of six diploid and one tetraploid species. *Plant J* 2013;75:699–709.
- [103] Gao D, Abernathy B, Rohksar D, Schmutz J, Jackson SA. Annotation and sequence diversity of transposable elements in common bean (*Phaseolus vulgaris*). *Front Plant Sci* 2014;5:339.
- [104] Vicent CM, Casacuberta JM. Additional ORFs in plant LTR-retrotransposons. *Front Plant Sci* 2020;11:555.
- [105] Orozco-Arias S, Isaza G, Guyot R. Retrotransposons in plant genomes: structure, identification, and classification through bioinformatics and machine learning. *Int J Mol Sci* 2019; 20:3837.
- [106] Almojil D, Bourgeois Y, Falis M, Hariyani I, Wilcox J, Boissinot S. The structural, functional and evolutionary impact of transposable elements in eukaryotes. *Genes* 2021;12:918.
- [107] Yuan YW, Wessler SR. The catalytic domain of all eukaryotic cut-and-paste transposase superfamilies. *Proc Natl Acad Sci U S A* 2011;108:7884–9.
- [108] Feschotte C, Pritham EJ. DNA transposons and the evolution of eukaryotic genomes. *Annu Rev Genet* 2007;41:331–68.
- [109] Xiong W, Dooner HK, Du C. Rolling-circle amplification of centromeric *Helitrons* in plant genomes. *Plant J* 2016;88:1038–45.
- [110] Kojima KK, Jurka J. *Crypton* transposons: identification of new diverse families and ancient domestication events. *Mob DNA* 2011;2:12.
- [111] Oliver KR, Greene WK. Transposable elements: powerful facilitators of evolution. *Bioessays* 2009;31:703–14.
- [112] Mhiri C, Borges F, Grandbastien MA. Specificities and dynamics of transposable elements in land plants. *Biology (Basel)* 2022; 11:488.
- [113] Chuong EB, Elde NC, Feschotte C. Regulatory activities of transposable elements: from conflicts to benefits. *Nat Rev Genet* 2017;18:71–86.
- [114] Kobayashi S, Goto-Yamamoto N, Hirochika H. Retrotransposon-induced mutations in grape skin color. *Science* 2004;304:982.
- [115] Dominguez M, Dugas E, Benchouaia M, Leduque B, Jimenez-Gomez JM, Colot V, et al. The impact of transposable elements on tomato diversity. *Nat Commun* 2020;11:4058.
- [116] Alonge M, Wang XG, Benoit M, Soyk S, Pereira L, Zhang L, et al. Major impacts of widespread structural variation on gene expression and crop improvement in tomato. *Cell* 2020;182: 145–61.e23.
- [117] Yu C, Zhang J, Peterson T. Genome rearrangements in maize induced by alternative transposition of reversed *Ac/Ds* termini. *Genetics* 2011;188:59–67.
- [118] Wang DF, Yu CH, Zuo T, Zhang JB, Weber DF, Peterson T. Alternative transposition generates new chimeric genes and segmental duplications at the maize *p1* locus. *Genetics* 2015; 201:925–35.
- [119] Gilbert C, Feschotte C. Horizontal acquisition of transposable elements and viral sequences: patterns and consequences. *Curr Opin Genet Dev* 2018;49:15–24.
- [120] Zhang H, Lang Z, Zhu JK. Dynamics and function of DNA methylation in plants. *Nat Rev Mol Cell Bio* 2018; 19:489–506.
- [121] Lei M, Zhang H, Julian R, Tang K, Xie S, Zhu JK. Regulatory link between DNA methylation and active demethylation in *Arabidopsis*. *Proc Natl Acad Sci U S A* 2015;112:3553–7.
- [122] Nystedt B, Street NR, Wetterbom A, Zuccolo A, Lin YC, Scofield DG, et al. The Norway spruce genome sequence and conifer genome evolution. *Nature* 2013;497:579–84.
- [123] Choi J, Lyons DB, Kim MY, Moore JD, Zilberman D. DNA methylation and histone H1 jointly repress transposable elements and aberrant intragenic transcripts. *Mol Cell* 2020;77: 310–23.e7.