Title: Patterns and Processes of Diploidization in Land Plants

Shortened Running Title: Diploidization in Land Plants

Authors (email; ORCID):

Zheng Li (liz7@email.arizona.edu; 0000-0001-6894-9616)

Michael T. W. McKibben (michaelmckibben@email.arizona.edu;

0000-0002-1342-0085)

Geoffrey S. Finch (gsfinch@email.arizona.edu; 0000-0002-5262-8719)

Paul D. Blischak (pblischak@email.arizona.edu; 0000-0001-9177-8958)

Brittany L. Sutherland (blsutherland@email.arizona.edu; 0000-0003-1773-6367)

Michael S. Barker (msbarker@arizona.edu; 0000-0001-7173-1319)

Corresponding Authors (email):

Zheng Li (liz7@email.arizona.edu)

Michael S. Barker (msbarker@arizona.edu)

Affiliation: Department of Ecology and Evolutionary Biology, University of Arizona,

Tucson, Arizona 85721, USA

Keywords

Diploidization, polyploidy, genome evolution, plant genomics, genome fractionation, chromosome pairing

Abstract

Most land plants are now known to be ancient polyploids that have rediploidized. This process of diploidization involves many changes in genome organization that ultimately restores bivalent chromosome pairing, disomic inheritance, and resolves dosage and other issues caused by genome duplication. Here, we provide an overview of the variety of mechanisms involved in diploidization as well as new analyses of pairing behavior and variation in gene fractionation across land plants. Overall, we find that lineage and WGD specific attributes influence the evolutionary outcomes of WGD and the process of diploidization in plant genomes. Ultimately, many of the mechanisms and forces driving diploidization remain to be discovered. Future research that leverages variation in the patterns and processes of diploidization will be able to advance our understanding of plant genome evolution and unlock the mysteries of diploidization.

INTRODUCTION

A major insight from two decades of sequencing plant genomes is that most are not simply diploid, but diploidized paleopolyploid genomes. Although it has long been

recognized that many contemporary plants are polyploids (7, 95, 140, 156), or species with duplicated genomes, it required comparative genomic analyses to provide conclusive evidence that plants experienced cycles of polyploidy followed by diploidization (4, 69, 80, 107, 147, 149, 153). Over the past century (9, 84, 152), we have learned a lot about polyploidization, but we know comparatively little about the mechanisms and forces that drive diploidization (35, 153). In the most basic sense, diploidization is the return of a polyploid genome to a diploid state (90, 138, 149, 153). One of the earliest references to this sort of diploidization—the fungal literature used "diploidization" in a different manner (eg: (48))—was by Stebbins (1947) in reference to a study by R. E. Clausen on pairing behavior in *Nicotiana* allopolyploids (27, 140, 141). The restoration of bivalent chromosome pairing behavior and associated diploid genetics is considered a key feature of diploidization. As recognized early on (141), the characteristics of a given whole genome duplication event impacts the pairing behavior, genetics, and subsequent course of diploidization in a polyploid genome. Thus, all polyploid species do not necessarily experience the same process of post-polyploid genome evolution and diploidization.

Although many mechanisms of genome evolution contribute to diploidization, it can be broadly described as involving two major processes: cytological diploidization and genic diploidization/fractionation (90, 93). Cytological diploidization occurs via chromosomal rearrangements, fission, fusion, and other large-scale chromosomal evolution events that produce significant changes in genome structure and eventually lead to diploid-like chromosome pairing behavior during meiosis (93). During

fractionation, many genes duplicated during the WGD event are lost, and only a subset of genes are retained as paralogs over time (47, 75). These two processes may occur largely independently of each other and at different rates yielding a diversity of genomes with different patterns of diploidization following polyploidy across lineages (89, 110, 153).

In this review, we discuss the different aspects of diploidization and post-polyploid genome evolution. We largely focus on genome evolution in the land plants, but also compare their patterns and processes of diploidization to those in animals and other eukaryotes. We begin with an introduction on the nature of polyploidy and how it may affect chromosome pairing behavior during meiosis. This includes a new survey of the plant cytological literature to assess the distribution of bivalent pairing among contemporary polyploid species. In the following sections we describe the two main processes of diploidization, cytological and genic diploidization, and summarize current knowledge on the molecular mechanisms of these diplodization processes. We also review differences in the rate of diploidization in plants and present new analyses on the rates of gene loss across land plants. Finally, we highlight the growing importance of developing new models and simulations to rigorously test hypotheses on diploidization as we try to understand the ultimate question: why diploidize at all?

THE NATURE OF POLYPLOIDY AND CHROMOSOME PAIRING BEHAVIOR

A key milestone during diploidization is establishing bivalent chromosome pairing during meiosis (153). Bivalent pairing is important because it is a precursor to restoring diploid-like genetics with two alleles per locus (ie., disomic inheritance). Although polyploids are often imagined to have multivalent pairing, many polyploid species actually have bivalent pairing at formation or evolve it quickly (145). Differences in pairing behavior are often used to distinguish the two major categories of polyploid species, allopolyploids and autopolyploids (7, 110, 120, 121). Distinguishing allo- and autopolyploids by pairing behavior is considered to be the “genetic classification” of polyploid species (7, 37, 38). In allopolyploids, divergence between the parental taxa is expected to limit pairing among the homoeologous chromosomes and the homologous chromosomes are expected to form pairs of bivalents during meiosis. In contrast, autopolyploids are expected to have homologous chromosomes that form multivalents (Figure 1). The bivalent pairing expected to occur in allopolyploids should lead to mostly disomic inheritance (two alleles at each of two distinct loci), whereas autopolyploids with multivalent pairing are expected to have multisomic inheritance (multiple alleles at a single locus) (Figure 1). It is important to point out that even though strictly bivalent pairing can occur in some autopolyploids, random segregation of homologous chromosomes during meiosis can result in multisomic inheritance (58, 67, 72, 119, 143). Therefore, multisomic inheritance is a unique feature to define autopolyploids (112, 145). Although the genetic definition is widely used in the field, many studies distinguish allo- and autopolyploid species by a taxonomic definition. This definition emphasizes the number of progenitor species (121). Allopolyploid species

result from hybridization of two or more species with genome duplication. In contrast, autopolyploids result from a genome duplication within a single progenitor species (7, 37). The taxonomic definition putatively gets around one of the limitations of the genetic definition: change in pairing behavior over time. As polyploid species diploidize, bivalent pairing is restored and this can make the genetic classification of an allo- or autopolyploid contingent on the age of the polyploid species. The taxonomic definition captures the nature of polyploid species regardless of the age of the WGD event and stage of diploidization.

Although the definitions of allo- and autopolyploidy are straightforward, in practice it is often difficult to describe the nature of polyploid species and degree of diploidization because of the dynamic processes of genome divergence and evolution. Allo- and autopolyploidy represent two ends of a continuum of variation in subgenome divergence and independence (7, 121, 141). This gradient of polyploid variation has long been recognized (140, 141). For example, the term “segmental allopolyploidy” was used for polyploid species that show mixtures of bivalent and multivalent formation (141). Differences in observed pairing behavior across this spectrum have been documented in multiple systems (121). This variation led to describing the inheritance patterns of segmental allopolyploids and other polyploids in the middle of this gradient of pairing behavior as being “mixosomic” (138). Although segmental allopolyploidy and mixosomic inheritance can be recognized by careful genetic analyses, most studies simply classify polyploid species as allo- or autopolyploids without distinguishing the polyploid variation continuum (7). However, to understand diploidization we ultimately

must grapple with this continuum of variation and recognize that not all studies of post-polyploid genome evolution are examining the same biology. For example, if a polyploid species is born with diploid-like bivalent pairing, is the ongoing divergent evolution of those homoeologous chromosomes really diploidization? Is it equivalent to the evolution of bivalent pairing in a multivalent autotetraploid? Analyses of diploidization in recent and ancient polyploid genomes need to better understand the origin of the species to evaluate what is and is not due to diploidization in these genomes.

One starting point to understanding diploidization in polyploid genomes is to assess how many contemporary polyploid species have bivalent pairing and how this pattern aligns with allo- and autopolyploid species. To address this gap in our knowledge, we conducted a survey of pairing behavior in allo- and autopolyploid species recognized by the taxonomic definition. The initial survey was based on a previous study of the frequency of allo- and autopolyploidy that examined data for 4,003 species from 47 genera of vascular plants (7). For each species, we recorded the chromosome pairing behavior during meiosis (Supplemental Table 1). We classified the meiotic chromosome pairing behavior as either strictly bivalent pairing (only bivalent formation was observed) or a mix (multivalent or mixture of bivalent and multivalent pairing). We identified 208 polyploid species from 40 genera (Supplemental Table 1) with at least one record of meiotic chromosome pairing behavior (Figure 2). Among these studies, 118 were classified by Barker et al. (7) as allopolyploids and 90 as autopolyploids (Figure 2, Supplemental Table 1). Overall, we found that 92 of these

species had strictly bivalent pairing, whereas 116 had mixed or multivalent pairing. Among species classified as allopolyploids, 48.3% had bivalent pairing and 51.7% had at least some multivalent formation during meiosis. Only 38.9% of the autopolyploids had bivalent pairing and 61.1% of the autopolyploids had multivalent or mixed pairing behavior. Consistent with our expectations, we found a lower frequency of strictly bivalent pairing among autopolyploid species compared to allopolyploids. However, the difference in pairing behavior between allo- and autopolyploids was not as large as expected. Some of this difference may be due to the taxonomic and phylogenetic classification of allo- and autopolyploid species used by Barker et al. (7), but the methodology used to classify polyploid species in that study is consistent with the approaches used broadly in the community. Our results suggest that segmental allopolyploidy is likely prevalent among polyploid plant species and that many autopolyploid species may rapidly evolve bivalent pairing.

Despite possessing twice the number of chromosomes as their progenitors and regardless of the nature of polyploid speciation, nearly half (44.2%) of the polyploid species we surveyed have bivalent chromosome pairing behavior. As expected, allopolyploid species demonstrated more strictly bivalent pairing than autopolyploid species. The stable meiosis of allopolyploid species likely results from pairing preferences for homologs and suppression of pairing between the divergent homoeologs (28, 110, 121). However, it has been suggested that stability of meiosis may be a neutral by-product of chromosomal divergence (63). Future studies need to determine whether and to what degree divergence among homoeologous

chromosomes leads to bivalent formation in polyploids. Further analyses on the divergence of the parental diploids and the pairing behavior of their allopolyploid species would provide some insight into this question. Similarly, analyses of the age of the surveyed autopolyploid species would help explain why nearly 40% had strictly 57 bivalent pairing. Are these species simply older autopolyploids that have gone through cytological diploidization already? Or are they cryptic allopolyploids that were misclassified as autopolyploids? The answers to these questions will help us understand the mechanisms that lead to the restoration of bivalent pairing in allo- and autopolyploids, and eventually the evolution of disomic inheritance across the spectrum of polyploid species.

MECHANISMS OF CYTOLOGICAL DIPLOIDIZATION

What are the mechanisms that lead to the restoration of bivalent pairing, disomic inheritance, and cytological diploidization of polyploid genomes in plants? Although the forces and mechanisms driving cytological diploidization are not completely understood (44, 63, 74), the process broadly involves changes in genome organization that ultimately produces pairs of homologous chromosomes that pair with each other and limit homoeologous pairing (Figure 3). These changes include chromosomal rearrangements, fissions, fusions, and other changes that lead to differentiated pairs of homologous chromosomes (74, 131). Dysploidy can also occur as a part of genome evolution associated with cytological diploidization, causing changes to base chromosome numbers (42, 90) and chromosome loss following WGD (86, 88, 91, 131,

159). More broadly, it is not yet clear if these changes accumulate (neutrally or through local adaptation) and lead to divergent resolution in different populations of a polyploid species (151), or if natural selection is driving cytological diploidization because of some fitness benefit of diploid genetics or meiosis.

Evidence from studies of established polyploid species indicates that natural selection is likely driving some aspects of cytological diploidization. Research on established polyploids suggests they have lower crossover frequencies compared to neotetraploids or their diploid relatives (121, 160). Recently formed polyploid species, especially autopolyploids but many allopolyploids as well (Figure 1), produce multivalents during meiosis. Multivalents are generally less stable during meiosis than bivalents and can lead to the loss of chromosomes during anaphase (74, 159, 163). This loss of chromosomes and other challenges of multivalent pairing and segregation can lead to reductions in fitness. These observations lead to a hypothesis that selection may reduce the number of crossovers or chiasma to suppress multivalent formation and non-homologous pairing in polyploid species (18, 25, 74). Reducing the number of crossovers limits the opportunity for chromosomes to pair with more than one partner during meiosis and leads to more stable, bivalent pairing.

In autopolyploids, meiotic stability is associated with the rate of crossover (17). More meiotically stable autopolyploids have diploid progenitors with a lower frequency of crossover formation, whereas polyploids with higher multivalent frequencies are formed by diploids with higher crossover rates (17, 61, 101). Studies suggest a single crossover per pair of homologous chromosomes is essential in most diploid species for

chromosome segregation (31, 70). For a chromosome to be associated with more than one partner during meiosis, at least two crossovers are required (17). Theoretically, reducing crossover to one per pair of homologous chromosomes in autopolyploids would be ideal for chromosome segregation and lead to bivalent formation (18). A model has been proposed for the mechanistic basis for limiting the number of crossovers in autopolyploids (18). In this model, the number of crossovers will be reduced to one if the range of crossover interference needs to be larger than the distance to the end of the chromosome (18). However, the genetic and molecular mechanisms that control the number of crossovers are not well understood. The genetic basis of autopolyploid meiosis has mainly been studied in autotetraploid *A. arenosa* (64, 100, 160). Previous studies used population data to show that eight unlinked candidate genes were important for meiotic chromosome pairing (64, 160). Strong signatures of selective sweeps are found on these genes and they are differentiated between polyploids and diploids. The results suggest that the genetics of re-establishing bivalent pairing in autopolyploid meiosis is likely to be polygenic (160). A more recent follow up study has identified the derived alleles of two genes, *ASY1* and *ASY3*, that are associated with meiotic changes in *A. arenosa* (100). This functional study also found that derived alleles of both genes are associated with traits in meiosis, such as reduction of multivalent formation, reduced chromosome axis length, and a tendency of more rod-shaped bivalent formation during meiosis (100). This work provides the first empirical analysis of multiple genes involved in bivalent restoration in autopolyploid meiosis and provides evidence that pairing behavior in autopolyploids

can be genetically controlled. Although this model of restoring bivalent pairing has been developed in the context of autopolyploid species, it likely applies to many allopolyploids that experience multivalent pairing as well (Figure 2).

Meiotic chromosome pairing behavior in allopolyploids is traditionally considered to be stable and diploid-like (28, 110, 121).The general explanation of the stable meiosis in allopolyploid species is that the homoeologous chromosomes are already differentiated, making it easier to establish bivalent pairing between homologs and suppress homoeolog pairing (28, 110, 121). The molecular mechanism that makes chromosome pairing behavior dependent on the divergence of chromosomes remains unclear (19, 29, 74). Further, many allopolyploid species still experience significant chromosomal change following genome duplication. Extensive chromosomal rearrangements and chromosome losses have been found in both synthetic *Brassica napus* and natural populations of *Tragopogon miscellus* (24, 159). As we found above (Figure 2), many allopolyploids also demonstrate some multisomic pairing and need to at least partially restore bivalent pairing to diploidize. Studies have shown that the restoration of diploid-like chromosome segregation is genetically controlled (44, 54, 63, 92, 125). The best known example is the *Ph1* locus, which has been studied in grasses, especially in wheat. This locus is associated with suppressing homoeologous pairing and promoting homologous chromosome pairing in meiosis. In the absence of *Ph1*, the number of crossovers increases and extensive homoeologous pairing can occur (127). Loci with similar effects have also been identified in allotetraploids *Brassica napus* (68, 79) and *A. suecica* (62). A recent study proposed a clear

mechanism of how non-homologous crossovers can be suppressed in allopolyploids (54). The gene *MSH4* is essential for the main crossover pathway in *B. napus*. The number of non-homologous crossovers decreases if *MSH4* returns to single copy and these crossovers will not be affected if *MSH4* is lost. Significantly, they found a convergent pattern of *MSH4* returning to a single copy following multiple independent WGDs across the angiosperms. However, researchers suggest *MSH4* is unlikely to contribute to meiosis stability in autopolyploids because it mainly affects non-homologous crossovers that are not thought to be important in autopolyploid pairing. This study provides a new mechanism for restoration of bivalent pairing in allopolyploids and suggests that chromosome pairing in allopolyploids is genetically determined across flowering plants (54).

Overall, the mechanisms behind restoring bivalent pairing is still not clear (44, 63, 74). Some evidence suggests chromosome pairing is genetically determined in different auto- and allopolyploid systems (62, 68, 79, 160). Few systems have been studied to understand the cytological diploidization of autopolyploids (19, 64, 100, 160). It remains unclear how these mechanisms may vary across the phylogeny. The recent study on *MSH4* shed some light on the molecular mechanism of cytological diploidization in flowering plants (54). Future studies should look for *MSH4* and other genes associated with pairing and test if chromosome pairing is genetically determined across land plants. The molecular mechanisms of cytological diploidization and the restoration of diploid-like bivalent pairing remain to be fully understood (19).

GENIC DIPLOIDIZATION AND FRACTIONATION

Although some polyploid species are essentially cytologically diploid at birth with bivalent pairing, all polyploid genomes appear to go through extensive gene loss and fractionation. Plant genomes are highly dynamic with significant turn-over in content, especially following WGDs (8, 129, 138, 149). All genes are duplicated during polyploidization and many of these new paralogs do not persist for long (1, 10, 30, 36, 134). This process of gene removal and loss following polyploidy is known as fractionation (47, 75). Although fractionation does not necessarily lead to the restoration of bivalent pairing or disomic inheritance, focussing on pairing behavior as the only process involved in diploidization misses the other aspects of genome evolution caused by WGDs. These include significant changes in gene content, network structure, and expression (16). Fractionation is a particularly important component of diploidization and post-polyploid genome evolution because they all experience gene loss and the resolution of duplicated gene networks.

Two major molecular mechanisms for fractionation have been proposed: pseudogenization and gene deletion by recombination (46, 47, 53). In flowering plants, it has been suggested that gene deletion by recombination is the predominant mechanism of fractionation and that pseudogenization may be relatively rare (47). However, a recent study estimated that the numbers of pseudogenes are highly lineage specific in angiosperm genomes, ranging from 5,000 to over 73,000 (158). These results suggest that pseudogenization may be more common in plant genomes than previously thought. Pseudogenization is generally caused by mutation and results in

the non-functionalization of a gene (157, 158, 165). Although gene function is lost, pseudogenes are not physically deleted from the genome. In contrast, gene deletion by recombination removes DNA from the genome (111, 155). Illegitimate recombination and unequal intra-strand homologous recombination are thought to be the two primary molecular mechanisms of gene deletion in plants (34, 139). These two mechanisms involve unequal crossing over during recombination and result in physical removal of DNA from the genome (34, 155). An additional potential mechanism for gene deletion in plants was recently proposed from research on synthetic allohexaploid *Brassica (49)*. Unlike the other two major molecular mechanisms of gene deletion, this deletion mechanism occurs between homoeologs during homoeologous recombination in allopolyploids. In the case of synthetic allohexaploid *Brassica*, fertility was significantly reduced when a particular subgenome was duplicated or deleted in a homoeologous exchange. This difference in fertility based on which subgenome is unbalanced in the homoeologous exchange can lead to a non-random retention of a subgenome (49).

In many other plant genomes, the process of fractionation has also been observed to be non-random (20, 52, 113, 148). This biased fractionation can result in subgenome dominance in which one subgenome is retained more than the other. This phenomenon has been widely observed across angiosperm lineages (23, 39, 41, 47, 122, 129). In general, genes from the more highly retained subgenome are expressed at a higher level than their homoeologs (23, 129). Transposable element (TE) density and methylation of these TEs can reduce the expression level of nearby genes (65, 66). In allopolyploids, one parental genome may have a higher TE density and higher

level of methylation compared to the other parental genome. It has been hypothesized that genes from the subgenome with higher TE density and methylation may be expressed at a lower level resulting in more fractionation compared to the other subgenome (154). Under this hypothesis, there is more opportunity for subgenome dominance to occur with allopolyploid species (154). This hypothesis has also been extended to paleopolyploidy events (50). It has been proposed that genomes with evidence of biased fractionation and subgenome dominance are more likely to be ancient allopolyploids (50). However, studies have shown that allopolyploid genomes may not always result in subgenome dominance. For example, in allopolyploids such as *B. napus*, wheat, and cotton, subgenome dominance is not observed (22, 57, 114, 161). In soybean, subgenome dominance is not found and the nature of its paleopolyploid event is still unresolved (166). These observations suggest the degree of genome differentiation prior to polyploidy may determine the amount of subgenome dominance. It remains unclear why this pattern varies across the phylogeny. Recent studies have provided progress on understanding the potential mechanisms that may drive subgenome dominance and biased fractionation. In the lotus genome, it has been found that subgenome dominance and biased fractionation is associated with higher gene body methylation, degree of protein-protein interactions, and gene expression levels (135). Recent studies also suggested homoeologous exchanges in allopolyploidy are likely to impact the pattern of subgenome dominance (3, 15, 40, 49). The phylogenetic distribution and relative contributions of these mechanisms to the evolution of subgenome dominance and biased fractionation is not yet clear, but

additional analyses leveraging population genomics, resynthesized polyploids, and other analyses of genetics and fitness will provide further insight into their roles in the polyploid genome evolution.

The drastic and biased gene loss that accompanies diploidization can also result in significant genome reorganization, which may occur to resolve genomic conflicts or dosage balance issues that would otherwise reduce polyploid fitness (116, 149). It has been shown that paralogs with more interaction partners, such as transcription factors, are more likely to be retained following WGD to maintain protein product stoichiometry or dosage (32, 45, 146). This dosage-balance hypothesis (DBH) also predicts that dosage-sensitive genes will be preferentially lost following small-scale gene duplication events to prevent dosage disruptions as their interaction partners are not doubled (13, 32, 45, 82). An alternative to the DBH attributes retention of paralogs to functional diversification, especially neofunctionalization (a gene copy acquiring a novel function) (106) or subfunctionalization (each gene copy retaining part of the original function) (85). A previous study suggests subfunctionalization may also drive cytological diploidization by maintaining appropriate chromosome pairs and promoting bivalent chromosome pairing and disomic inheritance (74). However, neo- and subfunctionalization cannot explain the parallel pattern of gene retention following different WGDs (10, 33, 89). Among these hypotheses for duplicate gene retention (45, 71), the DBH is the only hypothesis that explicitly predicts the parallel retention and loss of functionally related genes across species following WGD (30, 45, 146). A recent study of tandem duplicate genes in mammals suggests that the DBH might explain the

initial survival of these gene duplicates and neo- or subfunctionalization may be more important for the long term retention of paralogs (73). It remains to be understood what determines the portion of retained duplicate genes that are explained by the DBH, neo- and subfunctionalization, and other processes, and how this pattern varies across different lineages of plants.

In general, genic diploidization/fractionation occurs after all WGDs. Although the complete set of forces and mechanisms that drive fractionation are not yet understood, there is plenty of evidence that the process is generally not random with regard to the subgenomes and types of genes that are retained and lost (20, 52, 113, 148). Future studies should aim to better understand how much fractionation is determined by the nature of polyploidy or other factors such as level of methylation in parental genomes. We also need to understand how genic diploidization and fractionation contribute to resolving genomic conflicts or dosage balance issues. This will help improve our understanding of the fate of duplicate genes from the WGD. Given that diverse mechanisms and forces appear to drive fractionation, the processes of genic diploidization may vary considerably among lineages.

RATE OF DIPLOIDIZATION IN PLANTS

The process of diploidization involves many mechanisms and forces, and it is not yet clear how they operate in different lineages of plants. Most studies on genetic and cytological diploidization have focused on the angiosperms. In *Tragopogon*, it has been shown that the parallel pattern of gene loss and chromosomal rearrangements

can be established in only 40 generations (21). Similarly, Xiong et al. studied 10 generations of the resynthesized allopolyploid *Brassica napus* and found evidence for many chromosomal rearrangements and aneuploidies (159). Although there is evidence for rapid chromosomal evolution following polyploidy, a recent study demonstrated that the rate of diploidization following WGD can vary among related lineages (89). In 13 independent Brassicaceae mesopolyploidies, multiple species displayed different degrees of diploidization yielding a range of chromosome numbers and rearrangements across lineages. The different levels of diploidization are not clearly predicted by the age of these polyploidy events (89). More striking, in a recent cytological study of a Brassicaceae tribe largely endemic to Australia, different lineages descending from a common allopolyploid ancestor can have different rates of diploidization (91). The difference in rate is mainly driven by the number of chromosomal rearrangements observed in each species (91). Given that the rate of diploidization can vary dramatically in the descendants of a single WGD, the rate of diploidization likely varies across different lineages of flowering plants. However, it is not yet clear how much the rates of different aspects of diplodization vary across the land plant phylogeny and the forces driving these differences in rate.

Relatively little is known about diploidization outside of angiosperms. A recent study in *Sequoia* confirms that an autopolyploidization event occurred around 33 Ma (132). However, *Sequoia* has apparently maintained multivalent pairing since this paleopolyploidy (142), suggesting a slow diploidization process in comparison to flowering plants (132). Although debated (126, 167), genomic analyses have inferred at

least three other ancient WGDs in the gymnosperms (80, 81, 107). Other recent studies have found evidence of neopolyploidy in *Ginkgo* (136, 137) and *Junipus* (43). These ancient and recent WGDs provide opportunities to estimate the rate of genic and cytological diploidization in gymnosperms. Better understanding of diploidization in gymnosperms may provide a new angle to understand why polyploidy is relatively rare in most of the gymnosperms (2). Similar to the gymnosperms, diploidization remains to be studied in ferns. It has been hypothesized that ferns experienced multiple rounds of ancient WGDs without losing their chromosomes following WGDs (11, 59, 60). In contrast to the flowering plants, diploidization in the ferns has been hypothesized to be predominantly driven by gene silencing or pseudogenization rather than gene deletion (6, 59, 103, 104). A few studies have identified multiple, silenced copies of nuclear genes in putatively diploid homosporous fern genomes (96, 97, 115) and the active process of gene silencing without chromosome loss in a polyploid genome (51). However, the molecular mechanism of gene fractionation and the rate of diploidization in ferns is still unknown. Two heterosporous fern genomes have been published (77). However, these two genomes might experience different processes of diploidization compared to the homosporous ferns which have much higher average chromosome numbers. Similar to the gymnosperms and ferns, relatively little is known about diploidization in the other lineages of land plants. Future studies should estimate the patterns and processes of diploidization with chromosome level genome assemblies of these lineages, especially mosses, Lycopodiaceae, Isoetaceae, and the homosporous ferns where polyploidy seems to be prominent (107).

Estimating the rate of genic and cytological diploidization in plants can be challenging because the process occurs across large timescales and requires substantial genomic data. Additional phylogenetic and cytological analyses could be used to develop greater insight into the rate of cytological diploidization (Figure 4-5). Similarly, the rate of gene loss following polyploidy can be estimated from recent studies on the incidence of paleopolyploidy across the plant phylogeny. With genomic and transcriptomic data, the rate of duplicated gene loss in ancient polyploids can be estimated by comparing the fraction of paralogs in a genome derived from a WGD and the age of the WGD across multiple events and species. In general, studies have used synteny or duplicate gene age distribution analyses to infer duplicate genes derived from the polyploidy events (55, 118, 123). The relative age of a WGD can be estimated using the synonymous divergence (*Ks*) of the paralogs in the WGD peak from a *Ks* plot. By plotting the fraction of retained WGD paralogs in the genome (% paleologs) against the median paralog divergence for a WGD, we can obtain an estimate of the variation in the rate of genic diploidization following ancient WGDs.

Previous research has found that the fraction of genes retained from WGDs decreases exponentially over time in flowering plants (55, 118, 123). To estimate variation in the rate of gene loss across land plants, we analyzed land plant transcriptomic data of 815 species which are inferred to have at least one round of ancient polyploidy from the One Thousand Plant Transcriptome (1KP) project (107). These species were organized into five major lineages of land plants: bryophytes, lycophytes, ferns, gymnosperms, and angiosperms (Supplemental Table 2). We used

mixture modeling to identify genes retained from the most recent ancient WGD that each species experienced based on the WGD peak in the *Ks* plot (81). The paralog divergence of the WGD was estimated by the median *Ks* value of the WGD peak. We estimated the fraction of paleologs by using the total number of genes retained from an ancient WGD divided by the total number of unigenes in the transcriptome (Supplemental Table 2). We then plotted the fraction of paleologs with paralog divergence (*Ks*) of the WGD for each species (Figure 4). To infer if there was a significant trend in the data, we fit linear and exponential models to the distribution (Supplemental Table 3). Consistent with previous research (118, 123), we found a decrease in the fraction of retained paleologs over time in the angiosperms (Figure 4, Supplemental Table 3). We also observed higher variation in the fraction of retained paralogs among relatively young WGDs (lower *Ks* values) compared to older WGDs (higher *Ks* value). In contrast, we observed an increase in the fraction of paleologs over time in the gymnosperms (Figure 4, Supplemental Table 3). The bryophytes, lycophytes, and ferns did not have a significant increase or decrease in the fraction of retained WGD paralogs over time (Figure 4).

One issue with analyses of ancient polyploidy is that many taxa may be closely related and some taxa may share the same ancient duplication event. To test whether there is any phylogenetic signal for the fraction of retained paralogs and the relative age of the polyploidy, we used the phylosig function in phytools R package (124). We found evidence of significant phylogenetic signal for all categories except fractions of paleologs in the ferns and lycophytes. To address the potential impact of these closely

related species and phylogenetically shared WGDs on the observed relationship between WGD age and paleolog retention, we used phylogenetic independent contrasts (PIC) to account for the phylogenetic relatedness among lineages in our dataset. Specifically, we transformed raw values of the fraction of genes retained from each WGD and *Ks* value of a WGD and the phylogeny from the 1KP project using the pic function in the ape R package (117). Similar to the results above, our phylogenetically-corrected analyses did not recover a significant relationship between gene loss and the relative age of the WGD event in bryophytes, lycophytes, and ferns (Figure 5A-D, Supplemental Table 3). The significant positive relationship observed in the gymnosperms was not significant after taking phylogeny into account (Figure 5D, Supplemental Table 3). Our phylogenetically-corrected analyses recover a significant linear fit ($p < 0.001$, Adjusted R-squared = 0.09593, slope = -0.04506) and a significant exponential fit ($p < 0.001$, b = -0.2032) in angiosperms (Figure 5E, Supplemental Table 3). Similar to studies that did not take phylogeny into account (118, 123), we observed that paleologs were lost over time. We found that the relative age of the WGDs explains about 10% of the variation in the amount of gene loss in the linear model fit after PIC (Supplemental Table 3). Our study provides the first observation of the rate of gene loss in other lineages of land plants. Unlike flowering plants, the amount of gene loss from a WGD does not appear to be correlated with the relative age of the WGDs in these lineages. Our results suggest the dominant mechanism of fractionation may vary across land plants, and appears to be different in angiosperms compared to other land plants. Considering that the relative age of the WGD explained a relatively small

amount of the variation in gene loss in angiosperms, other mechanisms are clearly important. It may be that each WGD ultimately experiences different patterns of fractionation. Every post-WGD lineage experiences different demography, selection pressures, and other population genetic differences that could drive unique rates of gene loss. Variation in all of these dimensions likely contributes to the differences in the patterns of fractionation we observed across the land plant phylogeny.

Our results highlight that there is still much we do not understand about diploidization. Although other analyses also suggest that the rate of diploidization is likely to vary across the phylogeny of plants (91), it is not clear why we observed no relationship between the age of a WGD (as inferred by paralog divergence) and the fraction of retained paralogs for most clades of land plants. Future studies are needed to understand if the angiosperms have evolved novel mechanisms of gene fractionation distinct from those found in other land plants. Sample size in other lineages may contribute to some of the differences we observed, but the bryophytes, ferns, and gymnosperms were all represented by more than 50 species. Given the potential importance of eliminating genes after WGD (14, 32, 45, 146), the apparently efficient gene fractionation in angiosperms may be a part of their evolutionary success. Similarly, more comprehensive analyses of pseudogenization across land plants are needed to understand variation in gene loss among lineages. It also remains to be resolved how allo- and autopolyploidy influences the rate of gene loss and chromosomal evolution. Analyses leveraging comparative genomic approaches from emerging chromosome level gymnosperm and homosporous fern genomes will be

important to address why these rates of diploidization differ across land plants. Similarly, deeper analyses of populations and species descended from the same WGD are needed to understand the forces that drive diploidization. Our analyses and others (55, 118, 123) indicate that there is ample variation in the rates of diploidization to begin understanding these forces.

DIFFERENCES IN DIPLOIDIZATION BETWEEN PLANTS AND ANIMALS

Variation in the patterns and rates of diploidization is also evident between plants and animals. In angiosperms, most of the gene loss that occurs during fractionation is attributed to intrachromosomal recombination (47, 128, 144, 155). However, in animals many gene losses appear to be caused by pseudogenization (47). Vertebrate genomes do not seem to rapidly remove functionless nonrepetitive DNA, and pseudogenes can be carried for tens of millions of years (12, 76, 98, 130).

Patterns of gene loss following paleopolyploidy have been studied in many flowering plants such as *A. thaliana* (20), *Brassica* (20, 52, 113, 148), maize (20, 52, 113, 148), as well as more recent cotton allopolyploids (150). A general pattern that has been found across these flowering plant genomes is that most of the gene losses are due to illegitimate recombination rather than gene pseudogenization (47, 128, 144, 155). In maize, around 10% of the paleologs have been removed after a whole genome duplication that occurred around 12 million years ago. These paralogs were deleted by intrachromosomal recombination facilitated by direct repeats flanking the

gene or exons (155). In *Brassica rapa*, gene loss following the Brassiceae paleohexaploidy was driven by the same gene deletion mechanism (144).

In contrast to plant genomes with rapid gene deletion caused by intrachromosomal recombination, pseudogenization appears to be the major gene loss mechanism in vertebrates (12, 76, 98, 130). The most common type of pseudogenization occurs when a gene is disrupted by mutations and becomes unexpressed or non-functional (164). For example, all of the nearly 200 genes lost since humans diverged from chimpanzees are present as pseudogenes in our genome (130). Another excellent example of slow gene deletion in vertebrates comes from the recently sequenced rainbow trout genome (12). Analyses of the genome revealed an ancient WGD shared by the salmonid family. After nearly 100 million years of evolution, syntenic analyses found that the two subgenomes are still highly collinear. Nearly half of the protein-coding genes are retained in the genome, and most of the gene loss is due to pseudogenization. They also estimated that the average rate of gene inactivation is ~170 genes per million years (12). Similarly, carp experienced a WGD 8-18 MYa. Analyses of the common carp genome found a slow rate of gene loss with 92% of the paralogs from the polyploid event still retained in both copies (78). In *Xenopus* frogs, there is significant pseudogene accumulation following an allopolyploidy event that occurred 17-18 MYa. Comparable to rainbow trout, around 64% of paralogs from the WGD experienced gene loss by pseudogenization (133). Different from the patterns observed in flowering plants, few large scale gene deletions have been observed in animals. Most genes are deleted independently from

neighboring genes by single gene deletion (133). Notably, vertebrates represent all of the currently studied post-polyploid animal genomes. It is not clear if this pattern of gene deletion following WGDs is shared by all animals (12, 76).

The slow rate of gene removal in animals contrasts with the flowering plant-centric perspective that genes are rapidly deleted and genomes highly re-organized following WGDs. Slow gene deletion may impede the rate at which dosage balance problems are resolved following WGDs as well as reduce the rate of diploidization. The rapid gene deletion in flowering plants may allow them to resolve dosage balance problems much faster than animals. This hypothesis might help explain why polyploidy is rarer in animals compared to plants (87, 102, 108). Future studies should confirm if this pattern of gene deletion is shared by all animals. Recent genomic analyses revealed multiple paleopolyploidies in the ancestry of various invertebrate lineages, such as insects, horseshoe crabs, spiders, and molluscs (26, 56, 83, 105, 162). These ancient polyploids can be used to test if this pattern of gene deletion is shared by invertebrates. To test this hypothesis, one needs to assess the average rate of pseudogenization and gene deletion following polyploidy in animals and compare it to plants. Synteny analyses on high quality animal and plant genomes are needed to estimate the average rate of gene loss. Variation in the rates and mechanisms of diploidization will likely be found. For example, a recent study using 13 *Paramecium* genomes show a slower post-WGD gene loss rate compared to plants and vertebrates (55). Future studies are needed to further investigate the mechanisms and patterns of gene deletion following WGDs across eukaryotes.

FINAL THOUGHTS AND FUTURE DIRECTIONS

Diploidization involves a diversity of mechanisms to return polyploid genomes to an effectively diploid state. New comparative and population genomic data combined with cytogenetic and molecular biological approaches will continue to uncover the genetics and biology of the mechanisms involved in diploidization. Perhaps the most important next step in improving our understanding of diploidization is developing a more rigorous and objective framework for testing hypotheses about diploidization. Many studies of diploidization are largely descriptive. This is fair because we are still in the relatively early days of discovering ancient WGDs and their legacies in eukaryotic genomes. As we move forward and more data become available, we need to work towards more explicit hypothesis testing of diploidization. There has been progress in this area for some aspects of diploidization, such as hypotheses on subgenome dominance (15). Developing model and simulation based approaches to evaluate and test diploidization hypotheses would push the field forward. For example, model-based analyses of chromosomal evolution first introduced with chromEvol provided a new phylogenetic framework to test hypotheses of cytological evolution (94). Similar modeling and simulation approaches would permit researchers to more rigorously test hypotheses and develop more informed expectations about the outcomes of diploidization caused by different mechanisms and forces. Ultimately, the scale of data will demand more rigorous approaches as single genome analyses make way for phylogenomic and population genomic investigations.

More rigorous analyses of diploidization will also allow us to address perhaps the most interesting question about the entire process: why diploidize at all? Given the prevalence of diploidy among eukaryotes and the frequency of polyploid speciation in plants, we can deduce that polyploid species either diploidize or go extinct (4, 5). Why do polyploid species ultimately diploidize? It may be that bivalent pairing is inherently more stable than multivalent pairing and increases fitness. Perhaps bivalent pairing eventually leads to disomic inheritance and chromosomal differentiation by drift (74). Alternatively, diploidization may be driven to more efficiently purge deleterious substitutions in polyploid genomes (109). It may be that natural selection is more efficient in diploid genomes (99, 110) and selection in the environment, rather than the genome, drives diploidization. Model and simulation based analyses of these and other hypotheses would provide new ways to explicitly test the ultimate causes and drivers of diploidization. Coupling comparative genomic analyses and data with studies that are explicitly aimed at measuring fitness of the changes associated with diploidization are also needed. A challenge of studying diploidization is that many of the processes happen in that shadowy area of inference where the power of population genetics starts to fade but comparative phylogenetics may not be possible because of too few species. Moving forward, a combination of explicit models and simulations with data from carefully selected systems will help shine a light on the shadow of polyploidy.

LITERATURE CITED

1. Adams KL, Wendel JF. 2005. Polyploidy and genome evolution in plants. *Curr.*

Opin. Plant Biol. 8(2):135–41

2. Ahuja MR. 2005. Polyploidy in gymnosperms: revisited. *Silvae Genetica* 54:59–69
3. Alger EI, Edger PP. 2020. One subgenome to rule them all: underlying mechanisms of subgenome dominance. *Curr. Opin. Plant Biol.* 54:108–13
4. Arrigo N, Barker MS. 2012. Rarely successful polyploids and their legacy in plant genomes. *Curr. Opin. Plant Biol.* 15(2):140–46
5. Baduel P, Bray S, Vallejo-Marin M, Kolář F, Yant L. 2018. The “Polyploid Hop”: shifting challenges and opportunities over the evolutionary lifespan of genome duplications. *Front. Ecol. Evol.* 6:117
6. Barker MS. 2013. Karyotype and genome evolution in pteridophytes. In *Plant Genome Diversity* Volume 2
7. Barker MS, Arrigo N, Baniaga AE, Li Z, Levin DA. 2016. On the relative abundance of autopolyploids and allopolyploids. *New Phytol.* 210(2):391–98
8. Barker MS, Baute GJ, Liu S-L. 2012. Duplications and turnover in plant genomes. In *Plant Genome Diversity* Volume 1
9. Barker MS, Husband BC, Pires JC. 2016. Spreading Winge and flying high: The evolutionary importance of polyploidy after a century of study. *Am. J. Bot.* 103(7):1139–45
10. Barker MS, Kane NC, Matvienko M, Kozik A, Michelmore RW, et al. 2008. Multiple paleopolyploidizations during the evolution of the Compositae reveal parallel patterns of duplicate gene retention after millions of years. *Mol. Biol. Evol.* 25(11):2445–55

11. Barker MS, Wolf PG. 2010. Unfurling fern biology in the genomics age. *BioScience* 60:177–185
12. Berthelot C, Brunet F, Chalopin D, Juanchich A, Bernard M, et al. 2014. The rainbow trout genome provides novel insights into evolution after whole-genome duplication in vertebrates. *Nat. Commun.* 5:3657
13. Birchler JA, Veitia RA. 2011. Protein-protein and protein-DNA dosage balance and differential paralog transcription factor retention in polyploids. *Front. Plant Sci.* 2:64
14. Birchler JA, Veitia RA. 2012. Gene balance hypothesis: connecting issues of dosage sensitivity across biological disciplines. *Proc. Natl. Acad. Sci. U. S. A.* 109(37):14746–53
15. Bird KA, VanBuren R, Puzey JR, Edger PP. 2018. The causes and consequences of subgenome dominance in hybrids and recent polyploids. *New Phytol.* 220(1):87–93
16. Blischak PD, Mabry ME, Conant GC, Chris Pires J. 2016. Integrating networks, phylogenomics, and population genomics for the study of polyploidy. *Annu. Rev. Ecol. Evol. Syst.* 49:253-278
17. Bomblies K, Higgins JD, Yant L. 2015. Meiosis evolves: adaptation to external and internal environments. *New Phytol.* 208(2):306–323
18. Bomblies K, Jones G, Franklin C, Zickler D, Kleckner N. 2016. The challenge of evolving stable polyploidy: could an increase in “crossover interference distance” play a central role? *Chromosoma* 125:287–300.
19. Bomblies K, Madlung A. 2014. Polyploidy in the *Arabidopsis* genus. *Chromosome*

Res. 22(2):117–34

20. Bowers JE, Chapman BA, Rong J, Paterson AH. 2003. Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. *Nature* 422(6930):433–38
21. Buggs RJA, Chamala S, Wu W, Tate JA, Schnable PS, et al. 2012. Rapid, repeated, and clustered loss of duplicate genes in allopolyploid plant populations of independent origin. *Curr. Biol.* 22(3):248–52
22. Chalhoub B, Denoeud F, Liu S, Parkin IAP, Tang H, et al. 2014. Early allopolyploid evolution in the post-Neolithic *Brassica napus* oilseed genome. *Science* 345(6199):950–53
23. Cheng F, Wu J, Fang L, Sun S, Liu B, et al. 2012. Biased gene fractionation and dominant gene expression among the subgenomes of *Brassica rapa*. *PLoS One*. 7(5):e36442
24. Chester M, Gallagher JP, Symonds VV, da Silva AVC, Mavrodiev EV, et al. 2012. Extensive chromosomal variation in a recently formed natural allopolyploid species, *Tragopogon miscellus* (Asteraceae). *Proc. Natl. Acad. Sci. U. S. A.* 109(4):1176–1181
25. Cifuentes M, Grandont L, Moore G, Chèvre AM, Jenczewski E. 2010. Genetic regulation of meiosis in polyploid species: new insights into an old question. *New Phytol.* 186(1):29–36
26. Clarke TH, Garb JE, Hayashi CY, Arensburger P, Ayoub NA. 2015. Spider transcriptomes identify ancient large-scale gene duplication event potentially

important in silk gland evolution. *Genome Biol. Evol.* 7(7):1856–70

27. Clausen RE. 1941. Polyploidy in *Nicotiana. Am. Nat.* 75:291–306.
28. Comai L. 2005. The advantages and disadvantages of being polyploid. *Nat. Rev. Genet.* 6(11):836–46
29. Comai L, Tyagi AP, Lysak MA. 2003. FISH analysis of meiosis in *Arabidopsis* allopolyploids. *Chromosome Res.* 11(3):217–26
30. Conant GC, Birchler JA, Chris Pires J. 2014. Dosage, duplication, and diploidization: clarifying the interplay of multiple models for duplicate gene evolution over time. *Curr. Opin. Plant Biol.* 19:91–98
31. Crismani W, Mercier R. 2012. What limits meiotic crossovers? *Cell Cycle*. 11(19):3527–28
32. Defoort J, Van de Peer Y, Carretero-Paulet L. 2019. The evolution of gene duplicates in angiosperms and the impact of protein-protein interactions and the mechanism of duplication. *Genome Biol. Evol.* 11(8):2292–2305
33. De Smet R, Adams KL, Vandepoele K, Van Montagu MCE, Maere S, Van de Peer Y. 2013. Convergent gene loss following gene and genome duplications creates single-copy families in flowering plants. *Proc. Natl. Acad. Sci. U. S. A.* 110(8):2898–2903
34. Devos KM, Brown JKM, Bennetzen JL. 2002. Genome size reduction through illegitimate recombination counteracts genome expansion in *Arabidopsis*. *Genome Res.* 12(7):1075–79
35. Dodsworth S, Chase MW, Leitch AR. 2016. Is post-polyploidization diploidization

the key to the evolutionary success of angiosperms? *Bot. J. Linn. Soc.* 180(1):1–5

36. Douglas GM, Gos G, Steige KA, Salcedo A, Holm K, et al. 2015. Hybrid origins and the earliest stages of diploidization in the highly successful recent polyploid *Capsella bursa-pastoris*. *Proc. Natl. Acad. Sci. U. S. A.* 112(9):2806–11
37. Doyle JJ, Egan AN. 2010. Dating the origins of polyploidy events. *New Phytol.* 186(1):73–85
38. Doyle JJ, Sherman-Broyles S. 2017. Double trouble: taxonomy and definitions of polyploidy. *New Phytol.* 213(2):487–93
39. Edger PP, Poorten TJ, VanBuren R, Hardigan MA, Colle M, et al. 2019. Origin and evolution of the octoploid strawberry genome. *Nat. Genet.* 51(3):541–47
40. Edger PP, Smith R, McKain MR, Cooley AM, Vallejo-Marin M, et al. 2017. Subgenome dominance in an interspecific hybrid, synthetic allopolyploid, and a 140-year-old naturally established neo-allopolyploid monkeyflower. *Plant Cell* 29(9):2150–67.
41. Emery M, Willis MMS, Hao Y, Barry K, Oakgrove K, et al. 2018. Preferential retention of genes from one parental genome after polyploidy illustrates the nature and scope of the genomic conflicts induced by hybridization. *PLoS Genet.* 14(3):e1007267
42. Escudero M, Martín-Bravo S, Mayrose I, Fernández-Mazuecos M, Fiz-Palacios O, et al. 2014. Karyotypic changes through dysploidy persist longer over evolutionary time than polyploid changes. *PLoS One*. 9(1):e85266
43. Farhat P, Hidalgo O, Robert T, Siljak-Yakovlev S, Leitch IJ, et al. 2019. Polyploidy

in the conifer genus *Juniperus*: An unexpectedly high rate. *Front. Plant Sci.* 10:676

44. Feldman M, Levy AA. 2012. Genome evolution due to allopolyploidization in wheat. *Genetics*. 192(3):763–74
45. Freeling M, Michael F. 2009. Bias in plant gene content following different sorts of duplication: tandem, whole-genome, segmental, or by transposition. *Annu. Rev. Plant Biol.* 60(1):433–53
46. Freeling M, Scanlon MJ, Fowler JE. 2015. Fractionation and subfunctionalization following genome duplications: mechanisms that drive gene content and their consequences. *Curr. Opin. Genet. Dev.* 35:110–18
47. Freeling M, Woodhouse MR, Subramaniam S, Turco G, Lisch D, Schnable JC. 2012. Fractionation mutagenesis and similar consequences of mechanisms removing dispensable or less-expressed DNA in plants. *Curr. Opin. Plant Biol.* 15(2):131–39
48. Fulton IW. 1950. Unilateral nuclear migration and the Interactions of haploid mycelia in the fungus *Cyathus stercoreus*. *Proc. Natl. Acad. Sci. U. S. A.* 36(5):306-312
49. Gaebelein R, Schiessl SV, Samans B, Batley J, Mason AS. 2019. Inherited allelic variants and novel karyotype changes influence fertility and genome stability in *Brassica* allohexaploids. *New Phytol.* 223(2):965-78
50. Garsmeur O, Schnable JC, Almeida A, Jourda C, D'Hont A, Freeling M. 2014. Two evolutionarily distinct classes of paleopolyploidy. *Mol. Biol. Evol.* 31(2):448–54
51. Gastony GJ. 1991. Gene silencing in a polyploid homosporous fern:

paleopolyploidy revisited. *Proc. Natl. Acad. Sci. U. S. A.* 88(5):1602–5

52. Gaut BS, Doebley JF. 1997. DNA sequence evidence for the segmental allotetraploid origin of maize. *Proc. Natl. Acad. Sci. U. S. A.* 94(13):6809–14
53. Gaut BS, Wright SI, Rizzon C, Dvorak J, Anderson LK. 2007. Recombination: an underappreciated factor in the evolution of plant genomes. *Nat. Rev. Genet.* 8(1):77–84
54. Gonzalo A, Lucas M-O, Charpentier C, Sandmann G, Lloyd A, Jenczewski E. 2019. Reducing *MSH4* copy number prevents meiotic crossovers between non-homologous chromosomes in *Brassica napus*. *Nat. Commun.* 10(1):2354
55. Gout J-F, Johri P, Arnaiz O, Doak TG, Bhullar S, et al. 2019. Universal trends of post-duplication evolution revealed by the genomes of 13 *Paramecium* species sharing an ancestral whole-genome duplication. *bioRxiv*
56. Hallinan NM, Lindberg DR. 2011. Comparative analysis of chromosome counts infers three paleopolyploidies in the Mollusca. *Genome Biol. Evol.* 3:1150–63
57. Harper AL, Trick M, He Z, Clissold L, Fellgett A, et al. 2016. Genome distribution of differential homoeologue contributions to leaf gene expression in bread wheat. *Plant Biotechnol. J.* 14(5):1207–14
58. Hauber DP, Reeves A, Stack SM. 1999. Synapsis in a natural autotetraploid. *Genome* 42(5):936-949
59. Haufler CH. 1987. Electrophoresis is modifying our concepts of evolution in homosporous pteridophytes. *Am. J. Bot.* 74(6):953–66
60. Haufler CH, Soltis DE. 1986. Genetic evidence suggests that homosporous ferns

with high chromosome numbers are diploid. *Proc. Natl. Acad. Sci. U. S. A.* 83(12):4389–93

61. Hazarika MH, Rees H. 1967. Genotypic control of chromosome behaviour in rye X. Chromosome pairing and fertility in autotetraploids. *Heredity* 22:317–32
62. Henry IM, Dilkes BP, Tyagi A, Gao J, Christensen B, Comai L. 2014. The *BOY NAMED SUE* Quantitative trait locus confers increased meiotic stability to an adapted natural allopolyploid of *Arabidopsis*. *Plant Cell* 26(1):181-94
63. Hollister JD. 2015. Polyploidy: adaptation to the genomic environment. *New Phytol.* 205(3):1034–39
64. Hollister JD, Arnold BJ, Svedin E, Xue KS, Dilkes BP, Bomblies K. 2012. Genetic adaptation associated with genome-doubling in autotetraploid *Arabidopsis arenosa* *PLoS Genet.* 8(12):e1003093
65. Hollister JD, Gaut BS. 2009. Epigenetic silencing of transposable elements: a trade-off between reduced transposition and deleterious effects on neighboring gene expression. *Genome Res.* 19(8):1419–28
66. Hollister JD, Smith LM, Guo Y-L, Ott F, Weigel D, Gaut BS. 2011. Transposable elements and small RNAs contribute to gene expression divergence between *Arabidopsis thaliana* and *Arabidopsis lyrata*. *Proc. Natl. Acad. Sci. U. S. A.* 108(6):2322–27
67. Jackson RC, Jackson JW. 1996. Gene segregation in autotetraploids: prediction from meiotic configurations. *Am. J. Bot.* 83(6):673–8
68. Jenczewski E, Eber F, Grimaud A, Huet S, Lucas MO, et al. 2003. *PrBn*, a major

gene controlling homeologous pairing in oilseed rape (*Brassica napus*) haploids. *Genetics* 164(2):645–53

69. Jiao Y, Wickett NJ, Ayyampalayam S, Chanderbali AS, Landherr L, et al. 2011. Ancestral polyploidy in seed plants and angiosperms. *Nature* 473(7345):97–100
70. Jones GH, Franklin FCH. 2006. Meiotic crossing-over: obligation and interference *Cell* 126(2):246–8.
71. Kondrashov FA, Kondrashov AS. 2006. Role of selection in fixation of gene duplications. *J. Theor. Biol.* 239(2):141–51
72. Landergott U, Naciri Y, Jakob Schneller J, Holderegger R. 2006. Allelic configuration and polysomic inheritance of highly variable microsatellites in tetraploid gynodioecious *Thymus praecox* agg.. *Theor Appl. Genet.* 113:453–65
73. Lan X, Pritchard JK. 2016. Coregulation of tandem duplicate genes slows evolution of subfunctionalization in mammals. *Science* 352(6288):1009–13
74. Le Comber SC, Ainouche ML, Kovarik A, Leitch AR. 2010. Making a functional diploid: from polysomic to disomic inheritance. *New Phytol.* 186(1):113–22
75. Leitch AR, Leitch IJ. 2008. Genomic plasticity and the diversity of polyploid plants. *Science* 320(5875):481–83
76. Lien S, Koop BF, Sandve SR, Miller JR, Kent MP, et al. 2016. The Atlantic salmon genome provides insights into rediploidization. *Nature* 533:200–5
77. Li F-W, Brouwer P, Carretero-Paulet L, Cheng S, de Vries J, et al. 2018. Fern genomes elucidate land plant evolution and cyanobacterial symbioses. *Nat Plants*. 4(7):460–72

78. Li J-T, Hou G-Y, Kong X-F, Li C-Y, Zeng J-M, et al. 2015. The fate of recent duplicated genes following a fourth-round whole genome duplication in a tetraploid fish, common carp (*Cyprinus carpio*). *Sci. Rep.* 5:8199
79. Liu Z, Adamczyk K, Manzanares-Dauleux M, Eber F, Lucas M-O, et al. 2006. Mapping *PrBn* and other quantitative trait loci responsible for the control of homeologous chromosome pairing in oilseed rape (*Brassica napus* L.) haploids. *Genetics* 174(3):1583–96
80. Li Z, Baniaga AE, Sessa EB. 2015. Early genome duplications in conifers and other seed plants. *Sci. Adv.* 1(10):e1501084
81. Li Z, Barker MS. 2020. Inferring putative ancient whole-genome duplications in the 1000 Plants (1KP) initiative: access to gene family phylogenies and age distributions. *Gigascience* 9(2):giaa004
82. Li Z, Defoort J, Tasdighian S, Maere S, Van de Peer Y, De Smet R. 2016. Gene duplicability of core genes Is highly consistent across all angiosperms. *Plant Cell* 28(2):326–44
83. Li Z, Tiley GP, Galuska SR, Reardon CR, Kidder TI, et al. 2018. Multiple large-scale gene and genome duplications during the evolution of hexapods. *Proc. Natl. Acad. Sci. U. S. A.* 115(18):4713–18
84. Lutz AM. 1907. A preliminary note on the chromosomes of *Oenothera lamarckiana* and one of its mutants, *O. gigas*. *Science* 26(657):151–52
85. Lynch M, Force A. 2000. The probability of duplicate gene preservation by subfunctionalization. *Genetics* 154(1):459–73

86. Lysak MA. 2014. Live and let die: centromere loss during evolution of plant chromosomes. *New Phytol.* 203(4):1082–89
87. Mable BK. 2004. “Why polyploidy is rarer in animals than in plants”: myths and mechanisms. *Biol. J. Linn. Soc. Lond.* 82(4):453–66
88. Mandáková T, Joly S, Krzywinski M, Mummenhoff K, Lysak MA. 2010. Fast diploidization in close mesopolyploid relatives of *Arabidopsis*. *Plant Cell* 22(7):2277–90
89. Mandáková T, Li Z, Barker MS, Lysak MA. 2017. Diverse genome organization following 13 independent mesopolyploid events in Brassicaceae contrasts with convergent patterns of gene retention. *Plant J.* 91(1):3–21
90. Mandáková T, Lysak MA. 2018. Post-polyploid diploidization and diversification through dysploid changes. *Curr. Opin. Plant Biol.* 42:55–65
91. Mandáková T, Pouch M, Harmanová K, Zhan SH, Mayrose I, Lysak MA. 2017. Multispeed genome diploidization and diversification after an ancient allopolyploidization. *Mol. Ecol.* 26(22):6445–62
92. Martín AC, Shaw P, Phillips D, Reader S, Moore G. 2014. Licensing MLH1 sites for crossover during meiosis. *Nat. Commun.* 5:4580
93. Ma X-F, Gustafson JP. 2005. Genome evolution of allopolyploids: a process of cytological and genetic diploidization. *Cytogenet. Genome Res.* 109(1-3):236–49
94. Mayrose I, Barker MS, Otto SP. 2010. Probabilistic models of chromosome number evolution and the inference of polyploidy. *Syst. Biol.* 59(2):132–44
95. Mayrose I, Zhan SH, Rothfels CJ, Magnuson-Ford K, Barker MS, et al. 2011.

Recently formed polyploid plants diversify at lower rates. *Science* 333(6047):1257

96. McGrath JM, Hickok LG. 1999. Multiple ribosomal RNA gene loci in the genome of the homosporous fern *Ceratopteris richardii*. *Can. J. Bot.* 77(8):1199–1202
97. McGrath JM, Hickok LG, Pichersky E. 1994. Assessment of gene copy number in the homosporous ferns *Ceratopteris thalictroides* and *C. richardii* (Parkeriaceae) by restriction fragment length polymorphisms. *Plant Syst. Evol.* 189(3-4):203–10
98. Meyer A, Schartl M. 1999. Gene and genome duplications in vertebrates: the one-to-four (-to-eight in fish) rule and the evolution of novel gene functions. *Curr. Opin. Cell Biol.* 11(6):699–704
99. Monnahan P, Brandvain Y. 2019. The effect of autopolyploidy on population genetic signals of hard sweeps. *Biol. Lett.* 16:20190796
100. Morgan C, Zhang H, Henry CE, Franklin FCH, Bomblies K. 2020. Derived alleles of two axis proteins affect meiotic traits in autotetraploid *Arabidopsis arenosa*. *Proc. Natl. Acad. Sci. U. S. A.* 117 (16) 8980-88
101. Morrison JW, Rajhathy T. 1960. Chromosome behaviour in autotetraploid cereals and grasses. *Chromosoma* 11**:**297–309
102. Muller HJ. 1925. Why polyploidy is rarer in animals than in plants. *Am. Nat.* 59(663):346–53
103. Nakazato T, Barker MS, Rieseberg LH, Gastony GJ. 2008. Evolution of the nuclear genome of ferns and lycophytes. In *Biology and evolution of ferns and lycophytes*. Cambridge University Press
104. Nakazato T, Jung M-K, Housworth EA, Rieseberg LH, Gastony GJ. 2006.

Genetic map-based analysis of genome structure in the homosporous fern *Ceratopteris richardii*. *Genetics* 173(3):1585–97

105. Nossa CW, Havlak P, Yue J-X, Lv J, Vincent KY, et al. 2014. Joint assembly and genetic mapping of the Atlantic horseshoe crab genome reveals ancient whole genome duplication. *Gigascience* 3(1):2047-217X-3-9

106. Ohno S. 1970. *Evolution by gene duplication*, Springer Press

107. One Thousand Plant Transcriptomes Initiative. 2019. One thousand plant transcriptomes and the phylogenomics of green plants. *Nature* 574(7780):679–85

108. Orr HA. 1990. “Why Polyploidy is Rarer in Animals Than in Plants” Revisited. *Am. Nat.* 136(6):759–70

109. Otto SP. 2007. The evolutionary consequences of polyploidy. *Cell* 131(3):452–62

110. Otto SP, Whitton J. 2000. Polyploid incidence and evolution. *Annu. Rev. Genet.* 34:401–37

111. Pang E, Cao H, Zhang B, Lin K. 2015. Crop Genome Annotation: A Case Study for the *Brassica rapa* Genome. In *Compendium of Plant Genomes,* pp. 53–64

112. Parisod C, Holderegger R, Brochmann C. 2010. Evolutionary consequences of autopolyploidy. *New Phytol.* 186(1):5–17

113. Paterson AH, Xiyin W, Jingping L, Haibao T. 2012. Ancient and recent polyploidy in monocots. In *Polyploidy and Genome Evolution*, pp. 93–108

114. Pfeifer M, Kugler KG, Sandve SR, Zhan B, Rudi H, et al. 2014. Genome interplay in the grain transcriptome of hexaploid bread wheat. *Science* 345(6194):1250091

115. Pichersky E, Soltis D, Soltis P. 1990. Defective chlorophyll a/b-binding protein genes in the genome of a homosporous fern. *Proc. Natl. Acad. Sci. U. S. A.* 87(1):195–99

116. Pires JC, Conant GC. 2016. Robust yet fragile: expression noise, protein misfolding, and gene dosage in the evolution of genomes. *Annu. Rev. Genet.* 50(1):113–31

117. Popescu A-A, Huber KT, Paradis E. 2012. ape 3.0: New tools for distance-based phylogenetics and evolutionary analysis in R. *Bioinformatics* 28(11):1536–37

118. Qiao X, Li Q, Yin H, Qi K, Li L, et al. 2019. Gene duplication and evolution in recurring polyploidization–diploidization cycles in plants. *Genome Biol.* 20(1):1-23

119. Qu L, Hancock JF, Whallon JH. 1998. Evolution in an autopolyploid group displaying predominantly bivalent pairing at meiosis: genomic similarity of diploid *Vaccinium darrowi* and Autotetraploid *V. corymbosum* (Ericaceae). *Am. J. Bot.* 85(5):698-703

120. Ramsey J, Schemske DW. 1998. Pathways, mechanisms, and rates of polyploid formation in flowering plants. *Annu. Rev. Ecol. Systemat.* 29:467–501

121. Ramsey J, Schemske DW. 2002. Neopolyploidy in flowering plants. *Annu. Rev. Ecol. Systemat.* 33:589-639

122. Renny-Byfield S, Gong L, Gallagher JP, Wendel JF. 2015. Persistence of subgenomes in paleopolyploid cotton after 60 My of evolution. *Mol. Biol. Evol.* 32(4):1063–71

123. Ren R, Wang H, Guo C, Zhang N, Zeng L, et al. 2018. Widespread whole

genome duplications contribute to genome complexity and species diversity in angiosperms. *Mol. Plant*. 11(3):414–28

124. Revell LJ. 2012. phytools: an R package for phylogenetic comparative biology (and other things) *Methods Ecol. Evol.* 3(2):217-223

125. Riley R, Chapman V. 1958. Genetic control of the cytologically diploid behaviour of hexaploid wheat. *Nature* 182:713–715

126. Ruprecht C, Lohaus R, Vanneste K, Mutwil M, Nikoloski Z, et al. 2017. Revisiting ancestral polyploidy in plants. *Sci. Adv*. 3(7):e1603195

127. Sánchez-Morán E, Benavente E, Orellana J. 2001. Analysis of karyotypic stability of homoeologous-pairing (ph) mutants in allopolyploid wheats. *Chromosoma*. 110(5):371–77

128. Schnable JC, Freeling M, Lyons E. 2012. Genome-wide analysis of syntenic gene deletion in the grasses. *Genome Biol. Evol.* 4(3):265–77

129. Schnable JC, Springer NM, Freeling M. 2011. Differentiation of the maize subgenomes by genome dominance and both ancient and ongoing gene loss. *Proc. Natl. Acad. Sci. U. S. A.* 108(10):4069–74

130. Schrider DR, Costello JC, Hahn MW. 2009. All human-specific gene losses are present in the genome as pseudogenes. *J. Comput. Biol.* 16(10):1419–27

131. Schubert I, Lysak MA. 2011. Interpretation of karyotype evolution should consider chromosome structural constraints. *Trends Genet.* 27(6):207–16

132. Scott AD, Stenz NWM, Ingvarsson PK, Baum DA. 2016. Whole genome duplication in coast redwood (*Sequoia sempervirens*) and its implications for

explaining the rarity of polyploidy in conifers. *New Phytol.* 221(1):186-193

133. Session AM, Uno Y, Kwon T, Chapman JA, Toyoda A, et al. 2016. Genome evolution in the allotetraploid frog *Xenopus laevis*. *Nature* 538(7625):336–43

134. Shi T, Huang H, Barker MS. 2010. Ancient genome duplications during the evolution of kiwifruit (*Actinidia*) and related Ericales. *Ann. Bot.* 106(3):497–504

135. Shi T, Rahmani RS, Gugger PF, Wang M, Li H, et al. 2020. Distinct expression and methylation patterns for genes with different fates following a single whole-genome duplication in flowering plants. *Mol. Biol. Evol.* msaa105

136. Šmarda P, Horová L, Knápek O, Dieck H, Dieck M, et al. 2018. Multiple haploids, triploids, and tetraploids found in modern-day "living fossil" *Ginkgo biloba*. *Hortic. Res.* 5:55

137. Šmarda P, Vesel\`y P, Šmerda J, Bureš P, Knápek O, Chytrá M. 2016. Polyploidy in a "living fossil" *Ginkgo biloba*. *New Phytol.* 212(1):11-14

138. Soltis DE, Visger CJ, Marchant DB, Soltis PS. 2016. Polyploidy: Pitfalls and paths to a paradigm. *Am. J. Bot.* 103(7):1146–66

139. Soltis PS, Marchant DB, Van de Peer Y, Soltis DE. 2015. Polyploidy and genome evolution in plants. *Curr. Opin. Genet. Dev.* 35:119–25

140. Stebbins GL Jr. 1950. *Variation and evolution in plants*, Columbia University Press

141. Stebbins GL Jr. 1947. Types of polyploids; their classification and significance. *Adv. Genet.* 1:403–29

142. Stebbins GL Jr. 1948. The chromosomes and relationships of *Metasequoia* and

Sequoia. *Science* 108(2796):95–98

143. Stift M, Berenos C, Kuperus P, van Tienderen PH. 2008. Segregation models for disomic, tetrasomic and intermediate inheritance in tetraploids: a general procedure applied to *Rorippa* (yellow cress) microsatellite data. *Genetics* 179(4):2113–23

144. Tang H, Woodhouse MR, Cheng F, Schnable JC, Pedersen BS, et al. 2012. Altered patterns of fractionation and exon deletions in *Brassica rapa* support a two-step model of paleohexaploidy. *Genetics* 190(4):1563–74

145. Tayalé A, Parisod C. 2013. Natural pathways to polyploidy in plants and consequences for genome reorganization. *Cytogenet. Genome Res.* 140(2-4):79–96

146. Thomas BC, Pedersen B, Freeling M. 2006. Following tetraploidy in an *Arabidopsis* ancestor, genes were removed preferentially from one homeolog leaving clusters enriched in dose-sensitive genes. *Genome Res.* 16(7):934–46

147. Van de Peer Y, Mizrachi E, Marchal K. 2017. The evolutionary significance of polyploidy. *Nat. Rev. Genet.* 18(7):411–24

148. Wang X, Wang H, Wang J, Sun R, Wu J, et al. 2011. The genome of the mesopolyploid crop species *Brassica rapa*. *Nat. Genet.* 43(10):1035–39

149. Wendel JF. 2015. The wondrous cycles of polyploidy in plants. *Am. J. Bot.* 102(11):1753–56

150. Wendel JF, Flagel LE, Adams KL. 2012. Jeans, genes, and genomes: cotton as a model for studying polyploidy. In *Polyploidy and Genome Evolution*, pp. 181–207

151. Werth CR, Windham MD. 1991. A Model for divergent, allopatric speciation of polyploid pteridophytes resulting from silencing of duplicate-gene expression. *Am. Nat.* 137(4):515–26

152. Winge O. 1917. The chromosomes. Their numbers and general importance. *Compt. Rend. Trav. du Lab. de Carlsberg*. 13:131–75

153. Wolfe KH. 2001. Yesterday's polyploids and the mystery of diploidization. *Nat. Rev. Genet.* 2:333–41

154. Woodhouse MR, Cheng F, Pires JC, Lisch D, Freeling M, Wang X. 2014. Origin, inheritance, and gene regulatory consequences of genome dominance in polyploids. *Proc. Natl. Acad. Sci. U. S. A.* 111(14):5283–88

155. Woodhouse MR, Schnable JC, Pedersen BS, Lyons E, Lisch D, et al. 2010. Following tetraploidy in maize, a short deletion mechanism removed genes preferentially from one of the two homeologs. *PLoS Biol.* 8(6):e1000409

156. Wood TE, Takebayashi N, Barker MS, Mayrose I, Greenspoon PB, Rieseberg LH. 2009. The frequency of polyploid speciation in vascular plants. *Proc. Natl. Acad. Sci. U. S. A.* 106(33):13875–79

157. Xie J, Chen S, Xu W, Zhao Y, Zhang D. 2019. Origination and function of plant pseudogenes. *Plant Signal. Behav.* 14(8):1625698

158. Xie J, Li Y, Liu X, Zhao Y, Li B, et al. 2019. Evolutionary origins of pseudogenes and their association with regulatory sequences in plants. *Plant Cell* 31(3):563–78

159. Xiong Z, Gaeta RT, Pires JC. 2011. Homoeologous shuffling and chromosome compensation maintain genome balance in resynthesized allopolyploid *Brassica*

napus. *Proc. Natl. Acad. Sci. U. S. A.* 108(19):7908–13

160. Yant L, Hollister JD, Wright KM, Arnold BJ, Higgins JD, et al. 2013. Meiotic adaptation to genome duplication in *Arabidopsis arenosa*. *Curr. Biol.* 23(21):2151–56

161. Yoo M-J, Szadkowski E, Wendel JF. 2013. Homoeolog expression bias and expression level dominance in allopolyploid cotton. *Heredity* 110(2):171–80

162. Yoshida M-A, Ishikura Y, Moritaki T, Shoguchi E, Shimizu KK, et al. 2011. Genome structure analysis of molluscs revealed whole genome duplication and lineage specific repeat variation. *Gene* 483(1-2):63–71

163. Zhang H, Bian Y, Gou X, Zhu B, Xu C, et al. 2013. Persistent whole-chromosome aneuploidy is generally associated with nascent allohexaploid wheat. *Proc. Natl. Acad. Sci. U. S. A.* 110(9):3447–52

164. Zhang J. 2003. Evolution by gene duplication: an update. *Trends Ecol. Evol.* 18(6):292–298

165. Zhang Z, Harrison PM, Liu Y, Gerstein M. 2003. Millions of years of evolution preserved: a comprehensive catalog of the processed pseudogenes in the human genome. *Genome Res.* 13(12):2541–58

166. Zhao M, Zhang B, Lisch D, Ma J. 2017. Patterns and consequences of subgenome differentiation provide insights into the nature of paleopolyploidy in plants. *Plant Cell* 29(12):2974–94

167. Zwaenepoel A, Van de Peer Y. 2019. Inference of ancient whole-genome duplications and the evolution of gene duplication and loss rates. *Mol. Biol. Evol.*

36(7):1384–1404

ACKNOWLEDGMENTS

We thank Cristian Román-Palacios and Anthony E. Baniaga for helpful discussions. M. S. B. was supported by US National Science Foundation (NSF) grants IOS-1339156 and EF-1550838.

TERMS AND DEFINITIONS LIST:

Homologous Chromosomes (Homologs): A set of chromosomes that pair up during meiosis I; one is of maternal origin and the other of paternal origin.

Homoeologous Chromosomes (Homoeologs): A set of chromosomes in an allopolyploid that are derived from different parental species and have shared homology.

Disomic inheritance: Regular pairing and segregation of two chromosomes that produces two alleles at a locus.

Tetrasomic inheritance: Pairing and segregation of four chromosomes that produces four alleles at a single locus.

Multisomic/Polysomic inheritance: Combinations of chromosome pairing and segregation that yield more than two alleles at a locus.

Mixosomic inheritance: Combination of disomic and multisomic inheritances in a species.

Allopolyploidy: Polyploid species formed by interspecific hybridization and whole genome duplication. Generally considered to have pairs of homologous chromosomes from each parent that form bivalents during meiosis.

Autopolyploidy: Polyploid species with a single progenitor species and typically expected to have sets of homologous chromosomes that form multivalents during meiosis.

Segmental Allopolyploidy: Polyploid species with a mixture of bivalent and multivalent chromosome pairing.

Bivalent: A pair of homologous chromosomes aligned on the meiotic spindle during meiosis I.

Multivalent: Three or more homologous chromosomes aligned on the meiotic spindle during meiosis I.

Fractionation/Genic diploidization: The process of gene removal and loss following polyploidy by molecular mechanisms such as pseudogenization and gene deletion by recombination.

Cytological diploidization: The process of chromosomal evolution and restoration of bivalent pairing and disomic inheritance following polyploidy.

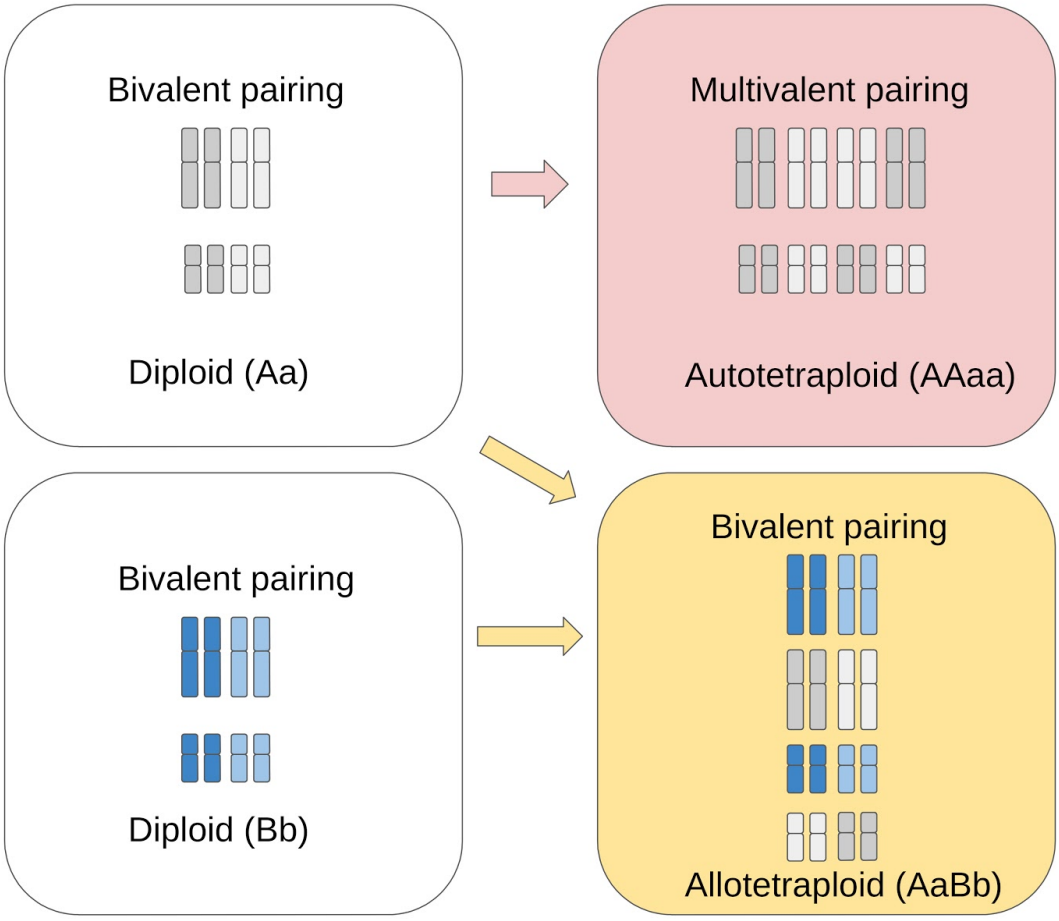

Figure 1. Chromosome pairing behavior during meiosis in diploid (white), autopolyploid (red), and allopolyploid (yellow). Chromosomes with the same size and color but different in shade represent homologous chromosomes. Chromosomes of the same size but in different colors (blue vs. gray) represent homoeologous chromosomes.

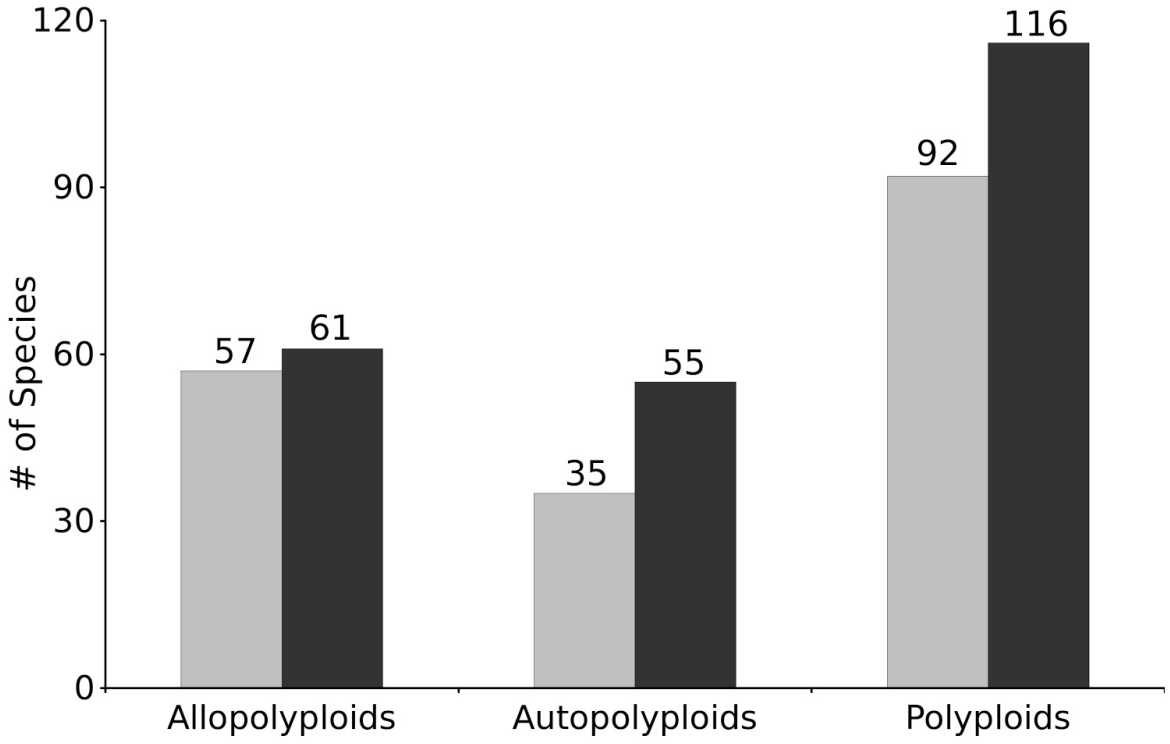

Figure 2. The frequency of strictly bivalent (gray) vs multivalent or a mix of bivalent and multivalent pairing (black) and whether species are reported as allo- or autopolyploids in Barker et al. (7). This meta-analysis is based on 208 species (Supplemental Table 1). The categories represent allopolyploids, autopolyploids, or all polyploids combined. The y-axis represents the number of species.

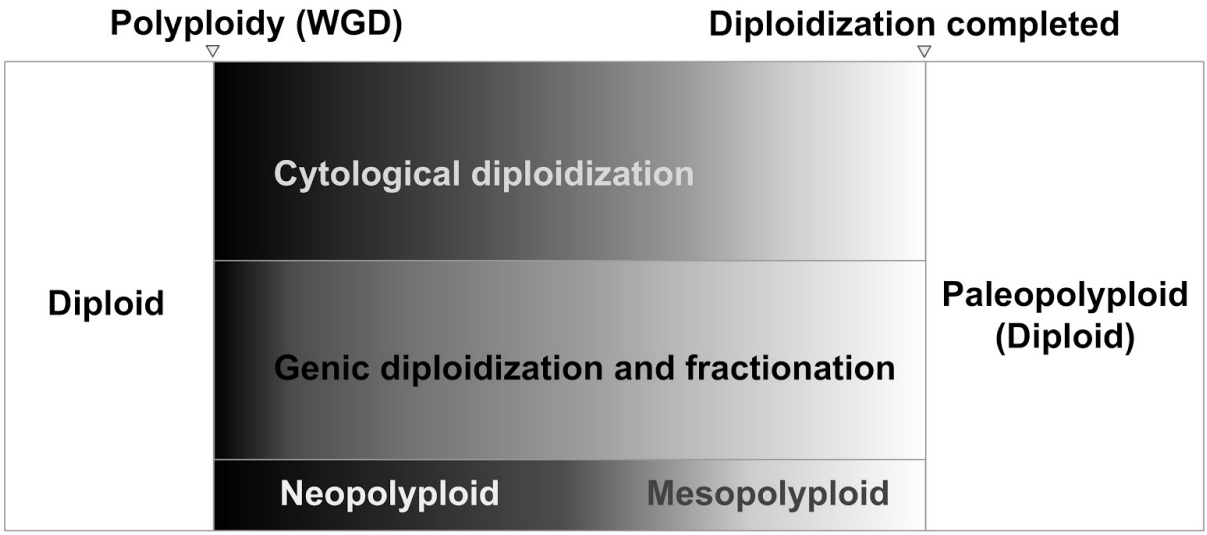

Figure 3. The major processes and mechanisms of diploidization. From left to right, the abrupt transition from white to black represents a change from diploidy to polyploidy. The gradual transition from gray to white represents diploidization. The shade of color shows the hypothetical level of diploidization. The differences in shade of color between cytological and genic diploidization shows that they are independent processes that occur at different rates. The process of cytological diploidization involves chromosomal evolution leading to the restoration of bivalent pairing and disomic inheritance following polyploidy. The process of genic diploidization and fractionation involves gene removal and loss following polyploidy by molecular mechanisms such as pseudogenization and gene deletion by recombination.

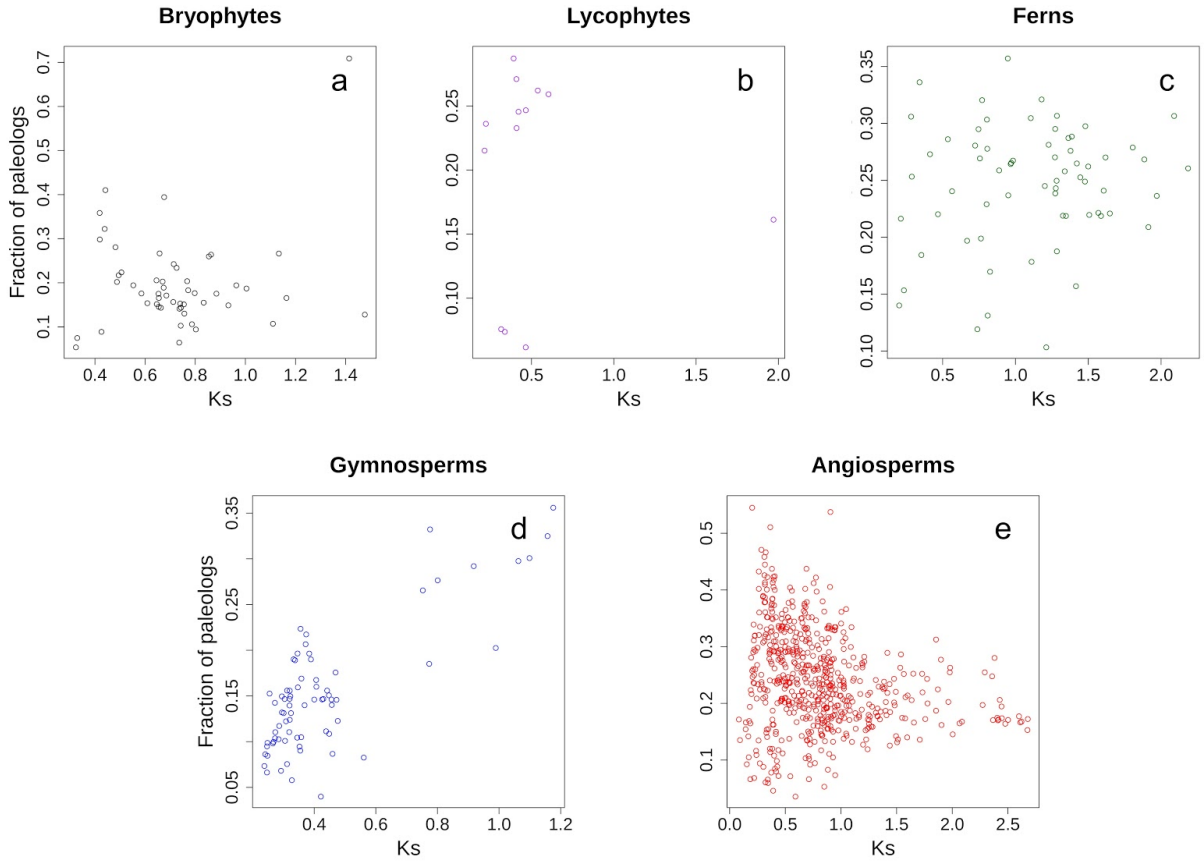

Figure 4. The fraction of genes retained from a WGD over estimated median *Ks* value of a WGD in land plants. The x-axis represents the *Ks* value of a WGD inferred by mixture model in gene age distribution analysis. The y-axis represents the fraction of gene retained from a WGD, which is estimated as the number of paralogs retained from a WGD divided by the total number of unigenes of a transcriptome. This study is based on 815 species of land plants (Supplemental Table 2; A: Bryophytes, 52 species; B: Lycophytes, 13 species; C: Ferns, 66 species, D: Gymnosperms, 73 species, E: Angiosperms, 610 species).

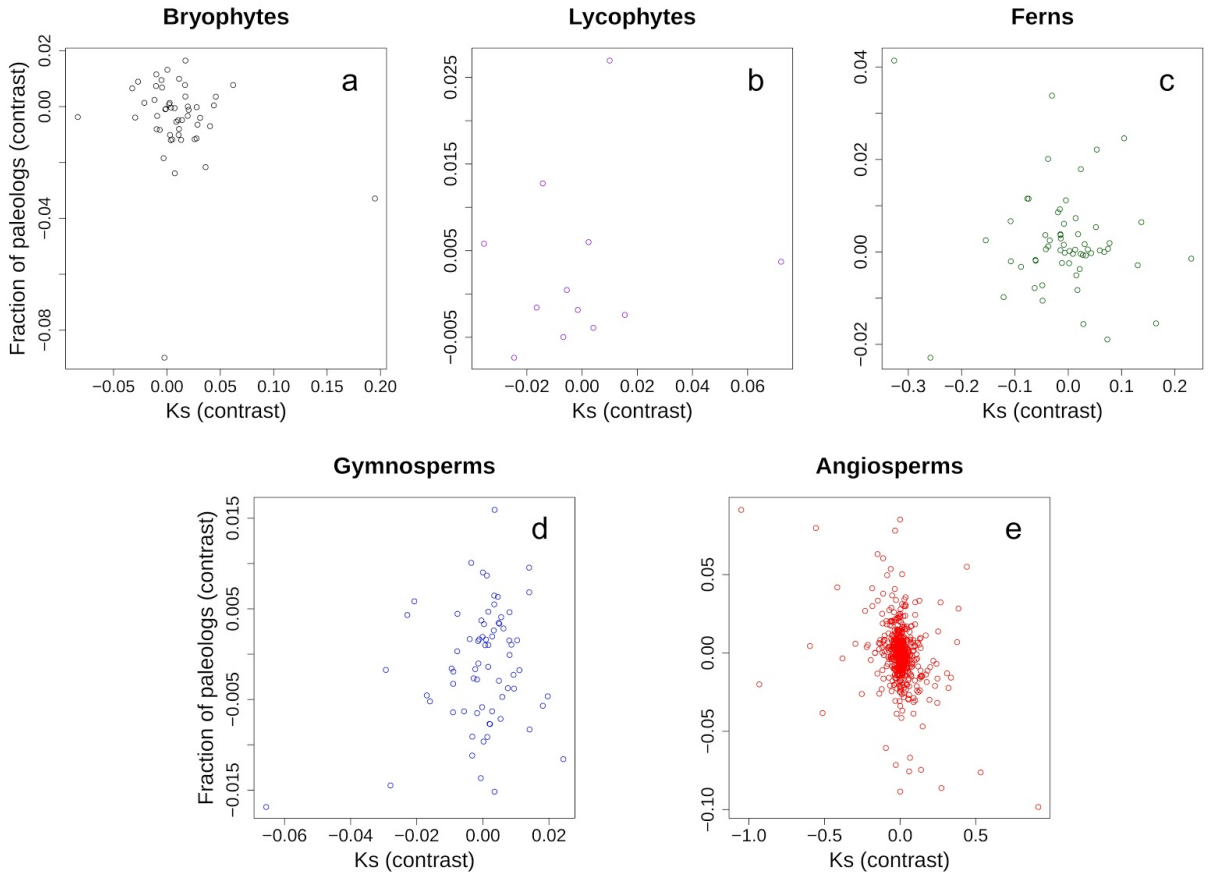

Figure 5. Phylogenetically-corrected rate of post-WGD paralog loss in land plants. Both the fraction of gene retained from a WGD (y-axis) and estimated median *Ks* value of a WGD (x-axis) in land plants were corrected using phylogenetic independent contrasts (PIC). This study is based on 815 species of land plants (Supplemental Table 2; A: Bryophytes, 52 species; B: Lycophytes, 13 species; C: Ferns, 66 species, D: Gymnosperms, 73 species, E: Angiosperms, 610 species).

SUPPLEMENTAL TABLES (available at https://doi.org/10.5281/zenodo.3963340)

Supplemental Table 1. Summary table of the frequency of strictly bivalent vs multivalent or a mix of bivalent and multivalent pairing in allo- or autopolyploids.

Supplemental Table 2. List of the fraction of genes retained from a WGD and the range and estimated median *Ks* value of each WGD analyzed.

Supplemental Table 3. Summary statistics of phylogenetic signal and linear and exponential fits before and after phylogenetically-corrected rate of post-WGD paralog loss in land plants.