CellPress
PARTNER JOURNAL

Molecular Plant
Review Article

Functional Divergence between Subgenomes and Gene Pairs after Whole Genome Duplications

Zhikai Liang and James C. Schnable*

Plant Science Innovation & Department of Agronomy and Horticulture, University of Nebraska-Lincoln, Lincoln, NE 68503, USA

*Correspondence: James C. Schnable (jschnable@unl.edu)

https://doi.org/10.1016/j.molp.2017.12.010

ABSTRACT

Gene loss following whole genome duplication (WGD) is often biased, with one subgenome retaining more ancestral genes and the other sustaining more gene deletions. While bias toward the greater expression of gene copies on one subgenome can explain bias in gene loss, this raises the question to what drives differences in gene expression levels between subgenomes. Differences in chromatin modifications and epigenetic markers between subgenomes in several model species are now being identified, providing an explanation for bias in gene expression between subgenomes. WGDs can be classified into duplications with higher, biased gene loss and bias in gene expression between subgenomes versus those with lower, unbiased rates of gene loss and an absence of detectable bias between subgenomes; however, the originally proposed link between these two classes and whether WGD results from an allo- or autopolyploid event is inconsistent with recent data from the allopolyploid *Capsella bursa-pastoris.* The gene balance hypothesis can explain bias in the functional categories of genes retained following WGD, the difference in gene loss rates between unbiased and biased WGDs, and how plant genomes have avoided being overrun with genes encoding dose-sensitive subunits of multiprotein complexes. Comparisons of gene expression patterns between retained transcription factor pairs in maize suggest the high degree of retention for WGD-derived pairs of transcription factors may instead be explained by the older duplication-degeneration-complementation model.

Key words: whole genome duplication, fractionation, subgenome, polyploidy

Liang Z. and Schnable J.C. (2018). Functional Divergence between Subgenomes and Gene Pairs after Whole Genome Duplications. Mol. Plant. **11**, 388–397.

INTRODUCTION

It has long been known that polyploidy is common among extant plant species, with estimates of the proportion of polyploid flowering ranging from 30% (Grant, 1963) to 80% (Masterson, 1994). Polyploids can be divided into two classes: autopolyploids, which carry more than two copies of the genome of a single species, and allopolyploids, which carry multiple copies of genomes from two or more distinct species. Over evolutionary time scales, the subgenomes of a polyploid can be rearranged through inversions, translocations, and chromosome fusions, masking the initial cytological marks of polyploidy. However, to date, large blocks of retained homeologous genes with conserved gene order (synteny) have been detected in almost every plant genome assembled to the pseudomolecule level, an indication of multiple rounds of whole genome duplication (WGD) resulting from polyploidy. Ancient polyploidy appears to be ubiquitous in the evolutionary history of flowering plants, with many lineages experiencing numerous reduplications (as reviewed by Wendel et al. [2016]) (Figure 1).

Genome doubling in polyploids can produce functionally distinct subgenomes. In some ancient polyploids, duplicate genes are lost preferentially from one copy of duplicated genomic segments (biased fractionation) (Thomas et al., 2006; Woodhouse et al., 2010; Cheng et al., 2012; Renny-Byfield et al., 2015). In the same ancient polyploids, expression tends to be biased between retained gene pairs (homeologs), with gene copies in regions where gene loss is less common tending to show greater mRNA abundance than the homeologs of those same genes in the corresponding genomic regions where gene loss is more common (Schnable et al., 2011b; Cheng et al., 2012; Garsmeur et al., 2013; Renny-Byfield et al., 2015). This pattern is referred to as genome dominance. While genome dominance, combined with the gene balance hypothesis, may explain biased fractionation (Figure 2), the origin of genome dominance, along with why some ancient polyploids fail to

Published by the Molecular Plant Shanghai Editorial Office in association with Cell Press, an imprint of Elsevier Inc., on behalf of CSPB and IPPE, SIBS, CAS.

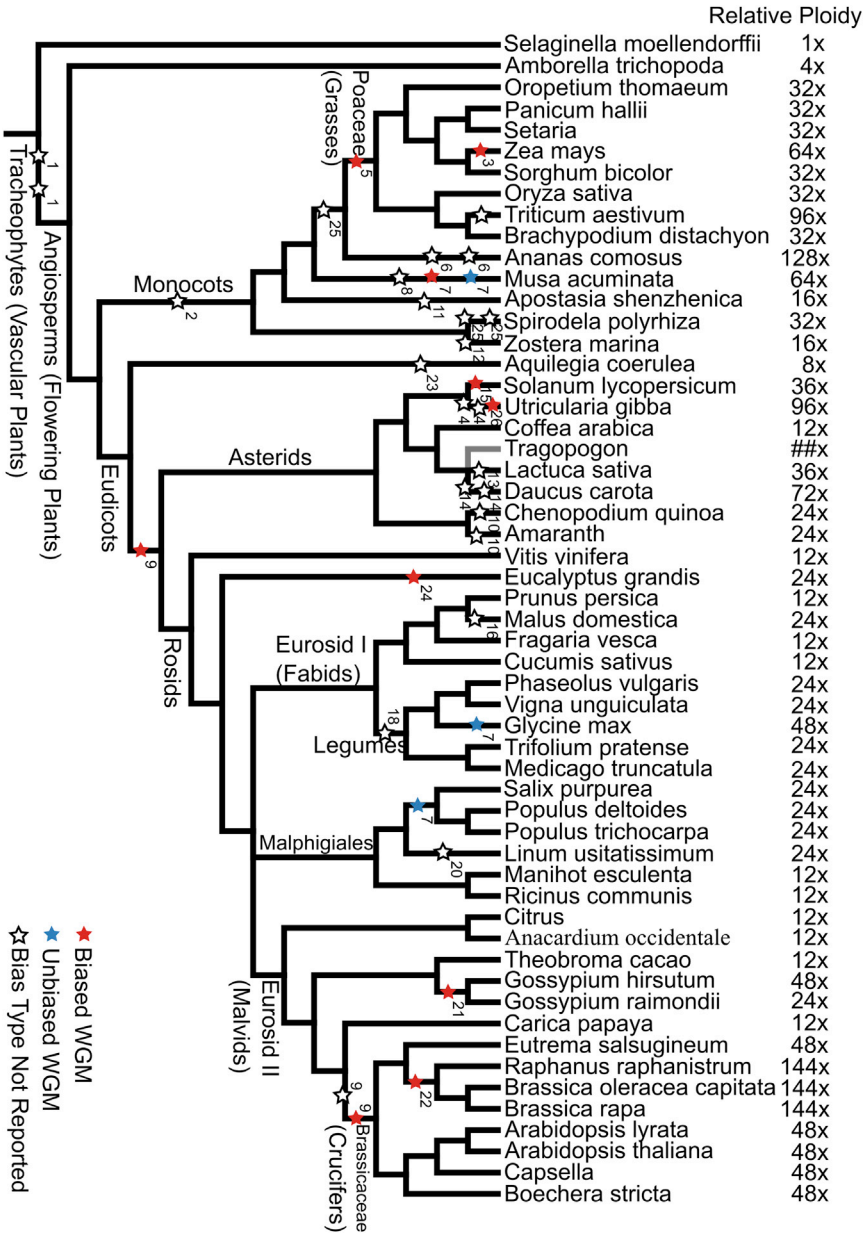

Figure 1. Placement and Bias WGDs Classification of Whole Genome Multiplications (Including Duplications, Triplications, and Complex or Unresolved Events) in the Flower Plant Lineage.

Subscript numerals next to each WGD on the phylogenetic tree indicate citations supporting the existence of that particular event. (1) Jiao et al. (2011); (2) Tang et al. (2010); (3) Schnable et al. (2011b); (4) Ibarra-Laclette et al. (2013); (5) Paterson et al. (2004); (6) Ming et al. (2015); (7) Garsmeur et al. (2013); (8) D'hont et al. (2012); (9) Bowers et al. (2003); (10) Lightfoot et al. (2017); (11) Zhang et al. (2017a); (12) Olsen et al. (2016); (13) Reyes-Chin-Wo et al. (2017); (14) Iorizzo et al. (2016); (15) Tomato Genome Consortium (2012); (16) Velasco et al. (2010); (17) Doyle and Egan (2010); (18) Cannon et al. (2010); (19) Tuskan et al. (2006); (20) Wang et al. (2012); (21) Li et al. (2015); (22) Moghe et al. (2014); (23) Tiley et al. (2016); (24) Myburg et al. (2014); (25) Wang et al. (2014); (26) Lan et al. (2017).

exhibit either genome dominance or biased fractionation, has remained unclear.

The predictions of the gene balance hypothesis are consistent with types of gene pairs that tend to be retained or lost following both WGDs and single gene duplications (Birchler and Veitia, 2012). However, recent evidence from expression analysis of transcription factors (a type of gene frequently retained as duplicate pairs following WGDs) suggests that the retention of this particular set of genes may be the result of the older duplication-degeneration-complementation model (Pophaly and Tellier, 2015). There is widespread interest in identifying cases of either subfunctionalization or neofunctionalization between homeologous gene pairs (Hughes et al., 2014; Li et al., 2016). Though duplicate genes arise through a range of mechanisms, homeologous gene pairs created through WGD are unique in that, in principle, both copies should initially be associated with the same regulatory sequences and chromatin environments. As expected, greater correlations in gene expression are observed between gene pairs resulting from WGD compared with duplicate genes of equivalent age resulting from other types of duplication (Casneuf et al., 2006; Wang et al., 2011b). Thus, differences in the pattern of gene-gene regulation between WGD-derived homeologous gene pairs should largely reflect changes in regulatory sequence or chromatin environment that occurred following WGD, a process sometimes referred to as fractionation mutagenesis (Freeling et al., 2012). However, divergence in regulation between orthologous genes in the parental species of an allopolyploid can confound such analysis (as reviewed by Buggs et al. [2014]). Initial reports suggest that variation in regulation patterns between different haplotypes within a single species, as well as orthologous genes in closely related species, can accumulate rapidly (Naito et al., 2009; Makarevitch et al., 2015; Waters et al., 2017; Zhang et al., 2017b), presenting significant challenges to studies of regulatory sub- or neofunctionalization in polyploids or ancient polyploids.

Definition Box

Homeology: a special case of paralogy describing the relationship between genes or genomic regions that diverged in a WGD event; these genes or genomic regions are referred to as "homeologous". Homeologous gene pairs are variously referred to in the literature as "homeologs", "syntenic paralogs", "ohnologs", or "paleologs". For further details on the origin of this term and how homeologous relationships are inferred, see Glover et al. (2016).

Co-orthologous: the shared relationship between two or more homeologous genes or genomic regions in a polyploid species to a single gene or genomic region in a diploid outgroup species.

Fractionation: the loss of redundant genes and/or noncoding regulatory elements from a genome following a whole-genome duplication.

Biased fractionation: a phenomenon observed in some, but not all, ancient WGDs where gene loss occurs unevenly between duplicated genomic regions, with one copy of a given region retaining many more ancestral genes than the other.

Subgenome: a set of genomic regions within a recent or ancient polyploid species derived from one parental species. A subgenome within a polyploid species is orthologous to the entire genome of diploid relatives; however, fractionation after polyploidy often results in significantly lower total gene content in subgenomes of polyploid species than in the genomes of diploid relatives.

Genome dominance: a bias toward greater transcript abundance of gene copies from one subgenome within a polyploid or ancient polyploid species compared with the homeologous copies of the same genes on another subgenome within the same species. The term originated in Flagel and Wendel (2010) and is used here, though the synonym "homeolog expression bias" has been proposed and is also used in the literature (Grover et al., 2012).

Ka/Ks: the ratio of the non synonymous substitution rate to synonymous substitution rate for a given protein coding sequence.

SUBGENOME-WIDE DIFFERENCES IN ANCIENT AND RECENT POLYPLOIDS

One of the first reported subgenome-wide patterns of divergence was bias in the pattern of gene loss following WGD (Thomas et al., 2006). WGD creates genetic redundancy for essentially all genes in the genome simultaneously, and in many cases one copy of the redundant gene pairs is lost from the genome through short-to medium-sized sequence deletions resulting from nonhomologous recombination (Woodhouse et al., 2010). In many species, this loss is biased between duplicated segments, with one region retaining a greater proportion of its ancestral gene content and the other region losing a greater proportion of its ancestral gene content (Thomas et al., 2006). While this observation was initially limited to individual duplicated segments of the genome (Thomas et al., 2006; Woodhouse et al., 2010), work in maize using a sorghum outgroup (Schnable et al., 2011b) and in *Brassica rapa* using an *Arabidopsis* outgroup (Wang et al., 2011a; Cheng et al., 2012) showed that this biased gene loss (biased fractionation) was consistent across entire reconstructed ancestral chromosomes, even after these sequences had been broken up and redistributed among the genome through inversions and translocations. Explaining this result requires either a model where fractionation is biased at a whole subgenome level, or where fractionation is biased at a whole-ancestral-chromosome level but where the ancestral subgenome that would be over- or under-fractionated is assigned stochastically on a chromosome-by-chromosome basis. While the former has generally been treated as the more parsimonious of the two explanations, it should be noted that the latter model has yet to be conclusively disproved.

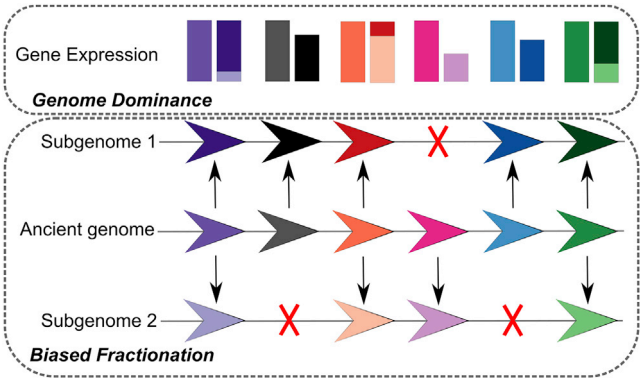

Figure 2. Biased Fractionation and Genome Dominance.
In the Biased Fractionation box, ancient and derived gene pairs are shown using the same color scheme, with arrows connecting homeologous gene pairs. The red Xs indicate genes lost through fractionation. In the Genome Dominance box, the similarly colored bars correspond with the genes in the Biased Fractionation box. The relative size of the colored bars in the Genome Dominance box indicates the relative amount of mRNA produced by each gene copy.

The use of biased fractionation as a mark enabled the provisional division of some ancient polyploid species into two or more putative subgenomes (Wang et al., 2011a; Schnable et al., 2011b; Tang et al., 2012; Murat et al., 2013; Renny-Byfield et al., 2015). Other functional differences have been shown to segregate with biased gene loss across the genome. The earliest of these was bias in expression between duplicate genes retained on multiple subgenomes. Bias in gene expression between retained duplicate genes on different subgenomes was reported in small sets of genes from recent allopolyploid cotton and resynthesized F1 crosses (Flagel and Wendel, 2010), along with the recent allopolyploid species *Arabidopsis suecica* (Chang et al., 2010). The correlation between genome dominance and biased fractionation was observed on a subgenome-wide scale in maize (Schnable et al., 2011b) and rapidly confirmed in *B rapa* (Cheng et al., 2012). The same correlation has since been reported in additional species, including cotton (Renny-Byfield et al., 2015). Recent support for the subgenome-wide model of bias in expression, and thus bias in gene loss, came from an analysis of resynthesized and natural monkeyflower *Mimulus peregrinus* populations that demonstrated that bias toward one parental subgenome was observable even in the first generation of the wide cross between the two parental species (*Mimulus guttatus* and *Mimulus luteus*), although the bias increased in subsequent generations (Edger et al., 2017).

While many ancient WGDs exhibit both biased gene loss and genome dominance, in several cases, including the most recent WGDs in poplar and soybean, little or no bias in patterns of gene loss or gene pair expression levels is observed between duplicated genomic segments (Garsmeur et al., 2013). These unbiased WGDs also retain many more duplicate gene pairs than WGDs of equivalent age in other lineages. The soybean (*Glycine max*) genome retains ~15 500 duplicate gene pairs from its most recent WGD (Schmutz et al., 2009), while the

maize genome, with a most recent WGD of equivalent age based on K_a/K_s ratios, retains ~4000 duplicate gene pairs (Schnable et al., 2009). The increased frequency of duplicate gene retention in polyploids that lack genome dominance is consistent with the gene balance hypothesis. More genes will show sensitivity to the loss of 50% of total gene product caused by the deletion of one copy of an equally expressed gene pair than to the loss of a minority of total gene product caused by the deletion of the less expressed copy of the gene pair where expression is biased.

When these two classes of ancient WGDs were identified, it was initially speculated that they represented the differences between duplication events resulting from allopolyploid and autopolyploid (Garsmeur et al., 2013). However, subsequent work has shown that recent allopolyploids can exhibit either pattern. The recent allopolyploid *Tragopogon miscellus* exhibits bias toward greater expressed duplicate gene copies from one parental diploid species (Buggs et al., 2010) and ~8% of a set of duplicate genes characterized fractionated within 40 generations (Buggs et al., 2012). The ~200 000-year-old allotetraploid *Capsella bursa-pastoris* shows evidence of neither significant bias toward a greater expression of gene copies of either diploid parent nor rates of gene loss from either subgenome that exceed the rates of gene loss in diploid *Capsella* species (Douglas et al., 2015). Thus, results from more recent allopolyploid species where both diploid progenitors are known and extant also identify two classes of WGD and recapitulate the link between genome dominance and biased fractionation; however, the reported lack of biased expression and biased and/or rapid gene loss in *Capsella* is inconsistent with biased WGDs originating from allopolyploidies and unbiased WGDs originating from autopolyploidies.

It must be noted that one of the most widely studied recent polyploids, hexaploid bread wheat (*Triticum aestivum*), has been inconsistently classified in the literature as either exhibiting genome dominance or not, perhaps a result of the special challenges faced by researchers trying to assign transcripts to homeologous gene copies when working with incomplete genome assemblies or genome assemblies from diploid or tetraploid relatives. Different studies have reported that expression is unbiased (Wang et al., 2017), biased toward greater expression of A and B homeologs at the expense of D genome homeologs (Li et al., 2014), and biased toward greater expression of the D genome (Leach et al., 2014). Other reports have found bias toward expression of gene copies from different subgenomes in different chromosome intervals (Harper et al., 2016). In contrast, gene loss data appear to clearly point to the D genome being dominant, with A intermediate and B experiencing the highest rate of gene loss (Pont et al., 2013). Overall subgenome-level analysis in wheat remains a complex challenge, although the continued improvement of the bread wheat genome assembly should address some of these issues (Zimin et al., 2017).

The observed correlation between biased gene loss and biased expression has spurred interest in identifying a mechanism, such as a chromatin mark, linking these two traits that might segregate between subgenomes. Consistent with the model of a chromatin mark being responsible for bias in gene expression between subgenomes, biased expression between retained duplicate genes was found to be reduced or abolished in pollen (Chettoor et al., 2014) where large-scale changes in chromatin also release many transposons from epigenetic silencing (Slotkin et al., 2009). Moderate differences in CG methylation were reported between subgenomes in *Brassica oleracea* (Parkin et al., 2014). The histone mark H3K4me3 has been reported to be significantly more common per megabase on the D subgenome of cotton than the A subgenome (Zheng et al., 2016), while the histone marks H3K4, H3K9, and H3K27me3 did not show significant bias between retained genes in the two maize subgenomes (Makarevitch et al., 2013; Renny-Byfield et al., 2017). Eichten et al. (2011) compared patterns of gene body methylation between retained duplicate genes in maize and found no significant differences between subgenomes; however, significant differences in methylation have been reported in upstream promoter sequences of genes retained in both maize subgenomes (Renny-Byfield et al., 2017) and *B rapa* subgenomes (Chen et al., 2015). Both repetitive sequence content and 24-nt small RNAs were found to be enriched on the non-dominant subgenomes of *B rapa* and cotton relative to retained copies of the same genes on the dominant subgenome (Woodhouse et al., 2014; Renny-Byfield et al., 2015; Cheng et al., 2016).

Genes on the dominant subgenome are under greater selective constraint as quantified by K_a/K_s ratios (Pophaly and Tellier, 2015; Yang et al., 2016). Gene pairs that exhibited biased expression toward one gene copy were twice as likely as other gene pairs to be targets of selection for domestication-related traits in *Brassica juncea* (Yang et al., 2016). This has led to speculation that the subordinate (less expressed) subgenome may have greater freedom to neofunctionalize and/or serve as a reservoir of phenotypic variation. Genes with retained duplicates from the maize WGD have been shown to be more frequently identified by GWAS hits in a meta-analysis of 41 different phenotypic datasets (Wallace et al., 2014), while genes on the dominant subgenome of maize have been shown to explain a greater proportion of total phenotypic variation than their retained duplicates on the subordinate subgenome (Renny-Byfield et al., 2017). A similar pattern of bias toward greater phenotypic effects for dominant subgenome genes was also observed for syntenic single-copy genes retained in only one subgenome (Renny-Byfield et al., 2017). In analyses of genes where one copy is retained and the other lost through fractionation, the use of an outgroup species is essential in separating retained syntenic genes from non-syntenic genes that were inserted into their present locations after the WGD being analyzed. Non-syntenic genes are much less likely to be assigned any Gene Ontology category annotations (Schnable et al., 2012a), tend to be transcribed at much lower levels (if at all), are significantly less likely to be translated into proteins when transcribed (Walley et al., 2016), and are much less likely to be the causal genes underlying mutations mapped and cloned using forward genetics (Schnable and Freeling, 2011).

CAUSES OF FRACTIONATION RESISTANCE among WGD GENE PAIRS

The genes retained as duplicate pairs or triplets following WGD are a nonrandom subset of the overall gene complement of the

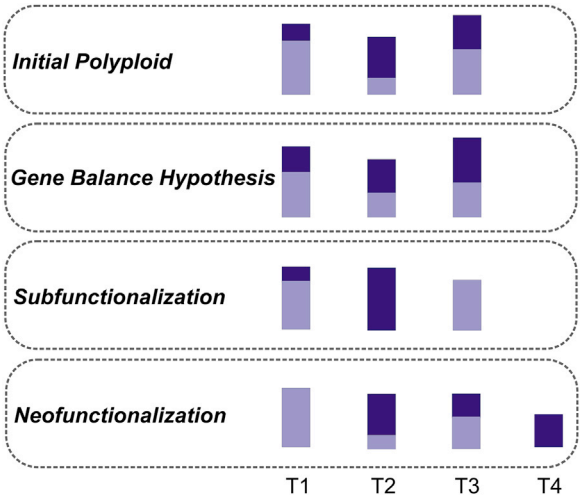

Figure 3. Predictions of Different Evolutionary Scenarios for the Expression Patterns for WGD Derived Duplicate Gene Pairs.
Predicted outcomes for the patterns of expression across four tissues for a duplicate gene pair retained following WGD as the result of three different mechanisms: gene balance constraint, subfunctionalization, and neofunctionalization. In each scenario the absolute height of the bar indicates the total expression of the gene pair in that tissue, while light and dark colors indicate the relative contribution of each gene copy to total gene pair expression in that tissue. In the ancestral state, both copies of the duplicated gene are initially expressed in three tissues (or cell types, or environmental contexts). Under the gene balance hypothesis, combined gene pair expression levels retain the same approximate ancestral levels. In both sub- and neofunctionalization, combined gene pair expression deviated from the ancestral state, and individual gene copies tend to be highly expressed in some tissues and expressed at low levels or not at all in others. Importantly, while the outcomes of sub- and neofunctionalization cannot be distinguished from each other without knowledge of the ancestral state, both produce different predicted outcomes from that of the gene balance hypothesis.

preduplication species. Genes retained as duplicate pairs following WGDs are disproportionately likely to encode transcription factors and subunits of multiprotein complexes (Blanc and Wolfe, 2004; Seoighe and Gehring, 2004; Maere et al., 2005). This pattern of biased gene pair retention has been explained as a consequence of the gene balance hypothesis, with WGD acting as a ratchet, enabling the duplication of otherwise dose-sensitive genes that are then unable to fractionate back to single-copy status without creating unfavorable changes in relative protein dosage (Birchler and Veitia, 2012). This model also predicts that genes retained following WGD should be under greater selective pressure to maintain constant levels of expression across different alleles within the same species and orthologous genes in related species. This prediction was recently validated across seven species in the genus *Glycine* (Coate et al., 2016). While repeated WGDs can result in long-term drive toward a greater abundance of certain types of genes in the genome over time (Freeling and Thomas, 2006), this is mitigated by mechanisms that allow repeatedly duplicated genes to escape from relative gene dosage constraints. Bias in gene expression level has been shown to be inherited across multiple sequential WGDs in grasses and crucifers, allowing the loss of one or both duplicate copies of gene pairs in the modern genome derived from a subordinate copy of a duplicate gene pair in an intermediately duplicated genome (Schnable et al., 2012b). In *Arabidopsis*, genes retained from a more recent WGD showed patterns of functional enrichment more consistent with the gene balance hypothesis than did sets of genes retained from an older WGD (Bekaert et al., 2011).

The gene balance hypothesis also predicts that, like gene copy number, gene expression pattern should be conserved between WGD-derived gene pairs. Analysis of expression patterns for retained duplicate gene pairs in modern species often shows significant divergence: less than 50% of retained *Arabidopsis* gene pairs from its most recent WGD showed significantly correlated expression patterns (Blanc and Wolfe, 2004). Similar results were reported by WGD-derived duplicate gene pairs in rice (Throude et al., 2009; Yim et al., 2009) and cotton (Renny-Byfield et al., 2014). Pophaly and Tellier (2015) used data from a set of maize reproductive tissues (Davidson et al., 2011) to demonstrate that retained WGD-derived duplicates in maize could be divided into two broad categories: gene pairs that retained largely correlated patterns of expression and genes with largely uncorrelated patterns of expression. The former category included many genes in macromolecular complexes, while the latter category was disproportionately enriched in transcription factors (Pophaly and Tellier, 2015). Conservation of correlated patterns of expression for both members of a homeologous gene pair is consistent with the retention of that gene pair as a result of dosage constraints. Uncorrelated expression across tissues would not support the conclusion that a particular gene pair was retained as a result of dosage constraints. The observations of (Pophaly and Tellier, 2015) therefore suggest that, unlike large protein complex subunits, other classes of genes, such as transcription factors, are preferentially retained following WGD for reasons other than dosage constraint, even though transcription factors are known to engage in protein-protein interactions. Transcription factors tend to be associated with large amounts of conserved regulatory sequence, creating the potential for regulatory subfunctionalization. Genes associated with larger amounts of conserved regulatory sequence are more likely to be retained as duplicate gene pairs following whole-genome duplication, even when controlling for the separate effects of protein function (Schnable et al., 2011a). The duplication, degeneration, complementation (DDC) model (Force et al., 1999; Lynch and Force, 2000) (Figure 3) predicts much less correlation between retained duplicate gene pairs than does the gene dosage balance hypothesis.

CIS-REGULATORY DIVERGENCE BEFORE AND AFTER WGD

One of the implicit assumptions in the interpretation of comparisons of WGD-derived duplicate pairs is that these genes start with equivalent complements of regulatory sequences as well as equivalent patterns of expression and diverge through loss-of-function mutations of specific enhancers or regulatory elements (Freeling et al., 2012). Examples of exceptions to this assumption of equivalent expression levels and patterns between orthologous genes in the diploid progenitor of allopolyploid species date back to early isozyme analysis of recent tetraploid species (Gottlieb, 1982; Ford and Gottlieb, 1999). Changes in the absolute level of gene expression

between different haplotypes in a single species (Hollister and Gaut, 2009) and orthologous genes in related species (Hollister et al., 2011) appear to be relatively common and may be the result of spreading of repressive chromatin marks from nearby transposon insertions (Hollister and Gaut, 2009; Hollister et al., 2011). The high degree of correlation in the expression pattern observed between duplicate genes in functional categories known to be dose sensitive suggests that, for these genes, patterns of transcriptional regulation are indeed subject to selective constraint (Pophaly and Tellier, 2015). However, work in both maize and *Capsella* has shown a high frequency of *cis*-regulatory variation across different alleles of the same genes in both species (Paschold et al., 2014; Josephs et al., 2015; Waters et al., 2017). A similar analysis comparing patterns of stress-responsive regulation for orthologous genes in maize and sorghum also concluded that the majority of stress-responsive patterns of gene regulation were lineage specific (Zhang et al., 2017b). These fast-accumulating changes in gene regulation may be the result of transposon insertions (Naito et al., 2009; Makarevitch et al., 2015; Steige et al., 2017). If many changes in gene regulation are neutral and fast-evolving character states driven by transposon insertions (Josephs et al., 2015; Zhang et al., 2017b), new gain-of-function regulatory variation may accumulate in each copy of duplicate gene pairs, resulting in incorrectly broad inferred ancestral patterns of expression. It is therefore likely that the sum of the tissue- and context-specific expression patterns of a duplicate gene pair is not a good estimator of the ancestral expression domain of the same gene prior to WGD.

Expression analysis of diploid progenitors and *in silico* combinations of the two can provide a point of comparison for the analysis of gene expression in allopolyploid species. Changes predicted by the *in silico* analysis are assumed to reflect divergence in gene regulation between the parental diploid species, while changes observed in the allopolyploid but not in the combined analysis of the diploid progenitors are considered to reflect either consequences of allopolyploidization itself or changes in *cis* or *trans* regulation that occurred after the genomes of the two diploid progenitors were combined into a single nucleus. This parents-plus-progeny approach assumes the direct diploid progenitors of a given allopolyploid are known and extant; however, even for many well-characterized recent allopolyploids, ''the diploids that we treat as 'parents' in our studies are not the actual parents, but closely related diploid lineages'' (Buggs et al., 2014). The problem of identifying true parents and not close relatives is exacerbated when substantial *cis*-regulatory variation is present and segregating within individual species (Josephs et al., 2015; Waters et al., 2017). Both divergence in gene regulation between related species and divergence in gene regulation between haplotypes in the same species will result in overestimating the degree of regulatory change associated with the formation and early evolution of polyploid species. One approach to reduce this overestimation would be to focus on only those patterns of gene regulation that appear to be under selective constraint through comparison of patterns of regulation for different haplotypes within a single species (Waters et al., 2017) or orthologous genes across multiple closely related species (Zhang et al., 2017b). While aspects of gene regulation subject to selective constraint may still diverge between orthologous genes in related species, these cases should be significantly less common, reducing the bias toward overestimating the regulatory changes associated with allopolyploidization.

The degree of observed correlation or divergence between WGD-derived duplicates may also depend on the types of expression data used. The majority of evidence for rapid divergence in *cis*-regulatory variation between haplotypes within the same diploid species or orthologous genes in related species uses data from abiotic stress treatments (Waters et al., 2017; Zhang et al., 2017b). The majority of evidence for gain-of-function changes in gene regulation resulting from transposon insertions also uses data from abiotic stress datasets (Naito et al., 2009; Makarevitch et al., 2015). An analysis in tetraploid *Capsella* identified the binding sites of the cold-responsive transcription factor CBF as one of the most common motifs appearing de novo in the proximal promoter of one member of a homeologous gene pair, but in neither orthologous gene copy in the diploid parental species (Kasianov et al., 2017). A comparison of regulation patterns across different floral tissues in grasses found an overall high degree of correlation in expression pattern between syntenic orthologous genes in related species (Davidson et al., 2012). Comparison of expression of syntenic orthologous genes in floral tissues of *Arabidopsis* and wild radish *Raphanus raphanistrum* (Figure 1) found that decreases or losses of gene expression by individual members of homeologous gene pairs or triplets were significantly more common than expression gains (Moghe et al., 2014), a result consistent with loss-of-function mutations in regulatory sequences being more common than gain-of-function mutations.

CONCLUSIONS AND FUTURE PERSPECTIVES

Although biased fractionation was first identified more than a decade ago, the processes responsible for creating biased fractionation and other associated phenomena, such as genome dominance, remain incompletely characterized. A range of chromatin, epigenetic, and transposon-associated marks have been shown to segregate between subgenomes where biased fractionation and genome dominance are observed (Zheng et al., 2016; Woodhouse et al., 2014; Renny-Byfield et al., 2015, 2017), but testing for causality between these correlations can be challenging. The existence of a second class of ancient WGD lacking biased fractionation, genome dominance, and with much slower rates of gene loss may prove a vital control case for testing epigenetic marks and other features thought to be associated specifically with biased fractionation and genome dominance. If data from *C. bursa-pastoris* indicating that this allopolyploid species is indeed unbiased are replicated in other recent allopolyploids, the conceptual model where biased ancient polyploids correspond with allopolyploids and unbiased ancient polyploids correspond with autopolyploids will be falsified. To the best of our knowledge, alternative conceptual models to explain why ancient polyploids appear to fall into distinct biased and unbiased categories have not been proposed, but alternative models to explain this phenomenon are clearly needed.

While the gene balance hypothesis continues to explain many traits of both recent and ancient polyploid species, it may be

the case that a subset of genes under complex regulatory control, particularly transcription factors, are retained through regulatory subfunctionalization consistent with the DDC model of duplicate gene evolution (Pophaly and Tellier, 2015). The study of regulatory subfunctionalization in polyploid species is complicated by the rapid emergence of *cis*-regulatory variation between haplotypes and orthologs (Naito et al., 2009; Makarevitch et al., 2015; Waters et al., 2017; Zhang et al., 2017b). Paired gene expression datasets from multiple outgroup species or large populations within a single species may significantly aid this area of investigation through the classification of features within the patterns of regulation for individual genes as functionally constrained or largely neutral.

FUNDING

This work was supported by the National Science Foundation under grant no. OIA-1557417 to J.C.S. and USDA NIFA award 2016-67013-24613 to J.C.S.

AUTHOR CONTRIBUTIONS

Z.L. and J.C.S. wrote the paper.

ACKNOWLEDGMENTS

The authors thank Xiaoxi Meng for helpful comments on an earlier draft of this manuscript. No conflict of interest is declared.

Received: September 17, 2017
Revised: November 28, 2017
Accepted: December 12, 2017
Published: December 22, 2017

REFERENCES

Bekaert, M., Edger, P.P., Pires, J.C., and Conant, G.C. (2011). Two-phase resolution of polyploidy in the *Arabidopsis* metabolic network gives rise to relative and absolute dosage constraints. Plant Cell **23**:1719–1728.

Birchler, J.A., and Veitia, R.A. (2012). Gene balance hypothesis: connecting issues of dosage sensitivity across biological disciplines. Proc. Natl. Acad. Sci. USA **109**:14746–14753.

Blanc, G., and Wolfe, K.H. (2004). Functional divergence of duplicated genes formed by polyploidy during *Arabidopsis* evolution. Plant Cell **16**:1679–1691.

Bowers, J.E., Chapman, B.A., Rong, J., and Paterson, A.H. (2003). Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. Nature **422**:433–438.

Buggs, R.J., Chamala, S., Wu, W., Gao, L., May, G.D., Schnable, P.S., Soltis, D.E., Soltis, P.S., and Barbazuk, W. (2010). Characterization of duplicate gene evolution in the recent natural allopolyploid *Tragopogon miscellus* by next-generation sequencing and Sequenom iPLEX MassARRAY genotyping. Mol. Ecol. **19** (*Suppl 1*):132–146.

Buggs, R.J., Chamala, S., Wu, W., Tate, J.A., Schnable, P.S., Soltis, D.E., Soltis, P.S., and Barbazuk, W.B. (2012). Rapid, repeated, and clustered loss of duplicate genes in allopolyploid plant populations of independent origin. Curr. Biol. **22**:248–252.

Buggs, R.J., Wendel, J.F., Doyle, J.J., Soltis, D.E., Soltis, P.S., and Coate, J.E. (2014). The legacy of diploid progenitors in allopolyploid gene expression patterns. Phil. Trans. R. Soc. Lond. B Biol. Sci. **369**:20130354.

Cannon, S.B., Ilut, D., Farmer, A.D., Maki, S.L., May, G.D., Singer, S.R., and Doyle, J.J. (2010). Polyploidy did not predate the evolution of nodulation in all legumes. PLoS One **5**:e11630.

Casneuf, T., De Bodt, S., Raes, J., Maere, S., and Van de Peer, Y. (2006). Nonrandom divergence of gene expression following gene and genome duplications in the flowering plant *Arabidopsis thaliana*. Genome Biol. **7**:R13.

Chang, P.L., Dilkes, B.P., McMahon, M., Comai, L., and Nuzhdin, S.V. (2010). Homeolog-specific retention and use in allotetraploid *Arabidopsis suecica* depends on parent of origin and network partners. Genome Biol. **11**:R125.

Chen, X., Ge, X., Wang, J., Tan, C., King, G.J., and Liu, K. (2015). Genome-wide DNA methylation pro-filing by modified reduced representation bisulfite sequencing in *Brassica rapa* suggests that epigenetic modifications play a key role in polyploid genome evolution. Front. Plant Sci. **6**:836.

Cheng, F., Wu, J., Fang, L., Sun, S., Liu, B., Lin, K., Bonnema, G., and Wang, X. (2012). Biased gene fractionation and dominant gene expression among the subgenomes of *Brassica rapa*. PLoS One **7**:e36442.

Cheng, F., Sun, C., Wu, J., Schnable, J., Woodhouse, M.R., Liang, J., Cai, C., Freeling, M., and Wang, X. (2016). Epigenetic regulation of subgenome dominance following whole genome triplication in *Brassica rapa*. New Phytol. **211**:288–299.

Chettoor, A.M., Givan, S.A., Cole, R.A., Coker, C.T., Unger-Wallace, E., Vejlupkova, Z., Vollbrecht, E., Fowler, J.E., and Evans, M.M. (2014). Discovery of novel transcripts and gametophytic functions via RNA-seq analysis of maize gametophytic transcriptomes. Genome Biol. **15**:414.

Coate, J.E., Song, M.J., Bombarely, A., and Doyle, J.J. (2016). Expression-level support for gene dosage sensitivity in three glycine subgenus glycine polyploids and their diploid progenitors. New Phytol. **212**:1083–1093.

D'hont, A., Denoeud, F., Aury, J.M., Baurens, F.C., Carreel, F., Garsmeur, O., Noel, B., Bocs, S., Droc, G., Rouard, M., et al. (2012). The banana (*Musa acuminata*) genome and the evolution of monocotyledonous plants. Nature **488**:213.

Davidson, R.M., Hansey, C.N., Gowda, M., Childs, K.L., Lin, H., Vaillancourt, B., Sekhon, R.S., de Leon, N., Kaeppler, S.M., Jiang, N., et al. (2011). Utility of RNA sequencing for analysis of maize reproductive transcriptomes. Plant Genome **4**:191–203.

Davidson, R.M., Gowda, M., Moghe, G., Lin, H., Vaillancourt, B., Shiu, S.H., Jiang, N., and Robin Buell, C. (2012). Comparative transcriptomics of three Poaceae species reveals patterns of gene expression evolution. Plant J. **71**:492–502.

Douglas, G.M., Gos, G., Steige, K.A., Salcedo, A., Holm, K., Josephs, E.B., Arunkumar, R., Ågren, J.A., Hazzouri, K.M., Wang, W., et al. (2015). Hybrid origins and the earliest stages of diploidization in the highly successful recent polyploid *Capsella bursa-pastoris*. Proc. Natl. Acad. Sci. USA **112**:2806–2811.

Doyle, J.J., and Egan, A.N. (2010). Dating the origins of polyploidy events. New Phytol. **186**:73–85.

Edger, P.P., Smith, R.D., McKain, M.R., Cooley, A.M., Vallejo-Marin, M., Yuan, Y.W., Bewick, A.J., Ji, L., Platts, A.E., Bowman, M.J., et al. (2017). Subgenome dominance in an interspecific hybrid, synthetic allopolyploid, and a 140-year-old naturally established neo-allopolyploid monkeyflower. Plant Cell **29**:2150–2167.

Eichten, S.R., Swanson-Wagner, R.A., Schnable, J.C., Waters, A.J., Hermanson, P.J., Liu, S., Yeh, C.T., Jia, Y., Gendler, K., Freeling, M., et al. (2011). Heritable epigenetic variation among maize inbreds. PLoS Genet. **7**:e1002372.

Flagel, L.E., and Wendel, J.F. (2010). Evolutionary rate variation, genomic dominance and duplicate gene expression evolution during allotetraploid cotton speciation. New Phytol. **186**:184–193.

Force, A., Lynch, M., Pickett, F.B., Amores, A., Yan, Y.L., and Postlethwait, J. (1999). Preservation of duplicate genes by complementary, degenerative mutations. Genetics **151**:1531–1545.

Ford, V., and Gottlieb, L. (1999). Molecular characterization of PgiC in a tetraploid plant and its diploid relatives. Evolution **53**:1060–1067.

Freeling, M., and Thomas, B.C. (2006). Gene-balanced duplications, like tetraploidy, provide predictable drive to increase morphological complexity. Genome Res. **16**:805–814.

Freeling, M., Woodhouse, M.R., Subramaniam, S., Turco, G., Lisch, D., and Schnable, J.C. (2012). Fractionation mutagenesis and similar consequences of mechanisms removing dispensable or less expressed DNA in plants. Curr. Opin. Plant Biol. **15**:131–139.

Garsmeur, O., Schnable, J.C., Almeida, A., Jourda, C., D'Hont, A., and Freeling, M. (2013). Two evolutionarily distinct classes of paleopolyploidy. Mol. Biol. Evol. **31**:448–454.

Glover, N.M., Redestig, H., and Dessimoz, C. (2016). Homeologs: what are they and how do we infer them? Trends Plant Sci. **21**:609–621.

Gottlieb, L.D. (1982). Conservation and duplication of isozymes in plants. Science **216**:373–380.

Grant, V. (1963). The Origin of Adaptations (New York, New York, USA: Columbia University Press).

Grover, C., Gallagher, J., Szadkowski, E., Yoo, M., Flagel, L., and Wendel, J. (2012). Homeolog expression bias and expression level dominance in allopolyploids. New Phytol. **196**:966–971.

Harper, A.L., Trick, M., He, Z., Clissold, L., Fellgett, A., Griffiths, S., and Bancroft, I. (2016). Genome distribution of differential homeologue contributions to leaf gene expression in bread wheat. Plant Biotechnol. J. **14**:1207–1214.

Hollister, J.D., and Gaut, B.S. (2009). Epigenetic silencing of transposable elements: a trade-off between reduced transposition and deleterious effects on neighboring gene expression. Genome Res. **19**:1419–1428.

Hollister, J.D., Smith, L.M., Guo, Y.L., Ott, F., Weigel, D., and Gaut, B.S. (2011). Transposable elements and small RNAs contribute to gene expression divergence between *Arabidopsis thaliana* and *Arabidopsis lyrata*. Proc. Natl. Acad. Sci. USA **108**:2322–2327.

Hughes, T.E., Langdale, J.A., and Kelly, S. (2014). The impact of widespread regulatory neofunctionalization on homeolog gene evolution following whole-genome duplication in maize. Genome Res. **24**:1348–1355.

Ibarra-Laclette, E., Lyons, E., Hernández-Guzmán, G., Pérez-Torres, C.A., Carretero-Paulet, L., Chang, T.H., Lan, T., Welch, A.J., Juárez, M.J., Simpson, J., et al. (2013). Architecture and evolution of a minute plant genome. Nature **498**:94–98.

Iorizzo, M., Ellison, S., Senalik, D., Zeng, P., Satapoomin, P., Huang, J., Bowman, M., Iovene, M., Sanseverino, W., Cavagnaro, P., et al. (2016). A high-quality carrot genome assembly provides new insights into carotenoid accumulation and asterid genome evolution. Nat. Genet. **48**:657–666.

Jiao, Y., Wickett, N.J., Ayyampalayam, S., Chanderbali, A.S., Landherr, L., Ralph, P.E., Tomsho, L.P., Hu, Y., Liang, H., Soltis, P.S., et al. (2011). Ancestral polyploidy in seed plants and angiosperms. Nature **473**:97–100.

Josephs, E.B., Lee, Y.W., Stinchcombe, J.R., and Wright, S.I. (2015). Association mapping reveals the role of purifying selection in the maintenance of genomic variation in gene expression. Proc. Natl. Acad. Sci. USA **112**:15390–15395.

Kasianov, A.S., Klepikova, A.V., Kulakovskiy, I.V., Gerasimov, E.S., Fedotova, A.V., Besedina, E.G., Kondrashov, A.S., Logacheva, M.D., and Penin, A.A. (2017). High quality genome assembly of *Capsella bursa-pastoris* reveals asymmetry of regulatory elements at early stages of polyploid genome evolution. Plant J. **91**:278–291.

Lan, T., Renner, T., Ibarra-Laclette, E., Farr, K.M., Chang, T.H., Cervantes-Pérez, S.A., Zheng, C., Sankoff, D., Tang, H., Purbojati, R.W., et al. (2017). Long-read sequencing uncovers the adaptive topography of a carnivorous plant genome. Proc. Natl. Acad. Sci. USA **114**:E4435–E4441.

Leach, L.J., Belfield, E.J., Jiang, C., Brown, C., Mithani, A., and Harberd, N.P. (2014). Patterns of homeologous gene expression shown by RNA sequencing in hexaploid bread wheat. BMC Genomics **15**:276.

Li, A., Liu, D., Wu, J., Zhao, X., Hao, M., Geng, S., Yan, J., Jiang, X., Zhang, L., Wu, J., et al. (2014). mRNA and small RNA transcriptomes reveal insights into dynamic homeolog regulation of allopolyploid heterosis in nascent hexaploid wheat. Plant Cell **26**:1878–1900.

Li, F., Fan, G., Lu, C., Xiao, G., Zou, C., Kohel, R.J., Ma, Z., Shang, H., Ma, X., Wu, J., et al. (2015). Genome sequence of cultivated upland cotton (*Gossypium hirsutum* TM-1) provides insights into genome evolution. Nat. Biotechnol. **33**:524–530.

Li, L., Briskine, R., Schaefer, R., Schnable, P.S., Myers, C.L., Flagel, L.E., Springer, N.M., and Muehlbauer, G.J. (2016). Co-expression network analysis of duplicate genes in maize (*Zea mays* L.) reveals no subgenome bias. BMC Genomics **17**:875.

Lightfoot, D., Jarvis, D., Ramaraj, T., Lee, R., Jellen, E., and Maughan, P. (2017). Single-molecule sequencing and Hi-C-based proximity-guided assembly of amaranth (*Amaranthus hypochondriacus*) chromosomes provide insights into genome evolution. BMC Biol. **15**:74.

Lynch, M., and Force, A. (2000). The probability of duplicate gene preservation by subfunctionalization. Genetics **154**:459–473.

Maere, S., De Bodt, S., Raes, J., Casneuf, T., Van Montagu, M., Kuiper, M., and Van de Peer, Y. (2005). Modeling gene and genome duplications in eukaryotes. Proc. Natl. Acad. Sci. USA **102**:5454–5459.

Makarevitch, I., Eichten, S.R., Briskine, R., Waters, A.J., Danilevskaya, O.N., Meeley, R.B., Myers, C.L., Vaughn, M.W., and Springer, N.M. (2013). Genomic distribution of maize facultative heterochromatin marked by trimethylation of H3K27. Plant Cell **25**:780–793.

Makarevitch, I., Waters, A.J., West, P.T., Stitzer, M., Hirsch, C.N., Ross-Ibarra, J., and Springer, N.M. (2015). Transposable elements contribute to activation of maize genes in response to abiotic stress. PLoS Genet. **11**:e1004915.

Masterson, J. (1994). Stomatal size in fossil plants: evidence for polyploidy in majority of angiosperms. Science **264**:421–424.

Ming, R., VanBuren, R., Wai, C.M., Tang, H., Schatz, M.C., Bowers, J.E., Lyons, E., Wang, M.L., Chen, J., Biggers, E., et al. (2015). The pineapple genome and the evolution of cam photosynthesis. Nat. Genet. **47**:1435–1442.

Moghe, G.D., Hufnagel, D.E., Tang, H., Xiao, Y., Dworkin, I., Town, C.D., Conner, J.K., and Shiu, S.H. (2014). Consequences of whole-genome triplication as revealed by comparative genomic analyses of the wild radish *Raphanus raphanistrum* and three other Brassicaceae species. Plant Cell **26**:1925–1937.

Murat, F., Zhang, R., Guizard, S., Flores, R., Armero, A., Pont, C., Steinbach, D., Quesneville, H., Cooke, R., and Salse, J. (2013). Shared subgenome dominance following polyploidization explains grass genome evolutionary plasticity from a seven protochromosome ancestor with 16k protogenes. Genome Biol. Evol. **6**:12–33.

Myburg, A.A., Grattapaglia, D., Tuskan, G.A., Vining, K., Amarasinghe, V., Ranik, M., Naithani, S., Elser, J., Liston, A., Spatafora, J.W., et al. (2014). The genome of *Eucalyptus grandis*. Nature **510**:356–362.

Naito, K., Zhang, F., Tsukiyama, T., Saito, H., Hancock, C.N., Richardson, A.O., Okumoto, Y., Tanisaka, T., and Wessler, S.R. (2009). Unexpected consequences of a sudden and massive

transposon amplification on rice gene expression. Nature **461**:1130–1134.

Olsen, J.L., Rouzé, P., Verhelst, B., Lin, Y.C., Bayer, T., Collen, J., Dattolo, E., De Paoli, E., Dittami, S., Maumus, F., et al. (2016). The genome of the seagrass *Zostera marina* reveals angiosperm adaptation to the sea. Nature **530**:331–335.

Parkin, I.A., Koh, C., Tang, H., Robinson, S.J., Kagale, S., Clarke, W.E., Town, C.D., Nixon, J., Krishnakumar, V., Bidwell, S.L., et al. (2014). Transcriptome and methylome profiling reveals relics of genome dominance in the mesopolyploid *Brassica oleracea*. Genome Biol. **15**:R77.

Paschold, A., Larson, N.B., Marcon, C., Schnable, J.C., Yeh, C.T., Lanz, C., Nettleton, D., Piepho, H.P., Schnable, P.S., and Hochholdinger, F. (2014). Nonsyntenic genes drive highly dynamic complementation of gene expression in maize hybrids. Plant Cell **26**:3939–3948.

Paterson, A., Bowers, J., and Chapman, B. (2004). Ancient polyploidization predating divergence of the cereals, and its consequences for comparative genomics. Proc. Natl. Acad. Sci. USA **101**:9903–9908.

Pont, C., Murat, F., Guizard, S., Flores, R., Foucrier, S., Bidet, Y., Quraishi, U.M., Alaux, M., Doležel, J., Fahima, T., et al. (2013). Wheat syntenome unveils new evidences of contrasted evolutionary plasticity between paleo-and neoduplicated subgenomes. Plant J. **76**:1030–1044.

Pophaly, S.D., and Tellier, A. (2015). Population level purifying selection and gene expression shape subgenome evolution in maize. Mol. Biol. Evol. **32**:3226–3235.

Renny-Byfield, S., Gallagher, J.P., Grover, C.E., Szadkowski, E., Page, J.T., Udall, J.A., Wang, X., Paterson, A.H., and Wendel, J.F. (2014). Ancient gene duplicates in *Gossypium* (cotton) exhibit near-complete expression divergence. Genome Biol. Evol. **6**:559–571.

Renny-Byfield, S., Gong, L., Gallagher, J.P., and Wendel, J.F. (2015). Persistence of subgenomes in paleopolyploid cotton after 60 my of evolution. Mol. Biol. Evol. **32**:1063–1071.

Renny-Byfield, S., Rodgers-Melnick, E., and Ross-Ibarra, J. (2017). Gene fractionation and function in the ancient subgenomes of maize. Mol. Biol. Evol. **34**:1825–1832.

Reyes-Chin-Wo, S., Wang, Z., Yang, X., Kozik, A., Arikit, S., Song, C., Xia, L., Froenicke, L., Lavelle, D.O., Truco, M.J., et al. (2017). Genome assembly with in vitro proximity ligation data and whole-genome triplication in lettuce. Nat. Commun. **8**:14953.

Schmutz, J., Cannon, S.B., Schlueter, J., Ma, J., Mitros, T., Nelson, W., Hyten, D.L., Song, Q., Thelen, J.J., Cheng, J., et al. (2009). Genome sequence of the palaeopolyploid soybean. Nature **463**:178–183.

Schnable, J.C., and Freeling, M. (2011). Genes identified by visible mutant phenotypes show increased bias toward one of two subgenomes of maize. PLoS One **6**:e17855.

Schnable, P.S., Ware, D., Fulton, R.S., Stein, J.C., Wei, F., Pasternak, S., Liang, C., Zhang, J., Fulton, L., Graves, T.A., et al. (2009). The B73 maize genome: complexity, diversity, and dynamics. Science **326**:1112–1115.

Schnable, J.C., Pedersen, B.S., Subramaniam, S., and Freeling, M. (2011a). Dose-sensitivity, conserved non-coding sequences, and duplicate gene retention through multiple tetraploidies in the grasses. Front. Plant Sci. **2**:2.

Schnable, J.C., Springer, N.M., and Freeling, M. (2011b). Differentiation of the maize subgenomes by genome dominance and both ancient and ongoing gene loss. Proc. Natl. Acad. Sci. USA **108**:4069–4074.

Schnable, J.C., Freeling, M., and Lyons, E. (2012a). Genome-wide analysis of syntenic gene deletion in the grasses. Genome Biol. Evol. **4**:265–277.

Schnable, J.C., Wang, X., Pires, J.C., and Freeling, M. (2012b). Escape from preferential retention following repeated whole genome duplications in plants. Front. Plant Sci. **3**:94.

Seoighe, C., and Gehring, C. (2004). Genome duplication led to highly selective expansion of the *Arabidopsis thaliana* proteome. Trends Genet. **20**:461–464.

Slotkin, R.K., Vaughn, M., Borges, F., Tanurdzić, M., Becker, J.D., Feijó, J.A., and Martienssen, R.A. (2009). Epigenetic reprogramming and small RNA silencing of transposable elements in pollen. Cell **136**:461–472.

Steige, K.A., Laenen, B., Reimegård, J., Scofield, D.G., and Slotte, T. (2017). Genomic analysis reveals major determinants of cis-regulatory variation in *Capsella grandiflora*. Proc. Natl. Acad. Sci. USA **114**:1087–1092.

Tang, H., Bowers, J.E., Wang, X., and Paterson, A.H. (2010). Angiosperm genome comparisons reveal early polyploidy in the monocot lineage. Proc. Natl. Acad. Sci. USA **107**:472–477.

Tang, H., Woodhouse, M.R., Cheng, F., Schnable, J.C., Pedersen, B.S., Conant, G., Wang, X., Freeling, M., and Pires, J.C. (2012). Altered patterns of fractionation and exon deletions in *Brassica rapa* support a two-step model of paleohexaploidy. Genetics **190**: 1563–1574.

Thomas, B.C., Pedersen, B., and 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**:934–946.

Throude, M., Bolot, S., Bosio, M., Pont, C., Sarda, X., Quraishi, U.M., Bourgis, F., Lessard, P., Rogowsky, P., Ghesquiere, A., et al. (2009). Structure and expression analysis of rice paleo duplications. Nucleic Acids Res. **37**:1248–1259.

Tiley, G.P., Ané, C., and Burleigh, J.G. (2016). Evaluating and characterizing ancient whole-genome duplications in plants with gene count data. Genome Biol. Evol. **8**:1023–1037.

Tomato Genome Consortium. (2012). The tomato genome sequence provides insights into fleshy fruit evolution. Nature **485**:635–641.

Tuskan, G.A., Difazio, S., Jansson, S., Bohlmann, J., Grigoriev, I., Hellsten, U., Putnam, N., Ralph, S., Rombauts, S., Salamov, A., et al. (2006). The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). Science **313**:1596–1604.

Velasco, R., Zharkikh, A., Affourtit, J., Dhingra, A., Cestaro, A., Kalyanaraman, A., Fontana, P., Bhatnagar, S.K., Troggio, M., Pruss, D., et al. (2010). The genome of the domesticated apple (*Malus* x *domestica* borkh.). Nat. Genet. **42**:833–839.

Wallace, J.G., Bradbury, P.J., Zhang, N., Gibon, Y., Stitt, M., and Buckler, E.S. (2014). Association mapping across numerous traits reveals patterns of functional variation in maize. PLoS Genet. **10**:e1004845.

Walley, J.W., Sartor, R.C., Shen, Z., Schmitz, R.J., Wu, K.J., Urich, M.A., Nery, J.R., Smith, L.G., Schnable, J.C., Ecker, J.R., et al. (2016). Integration of omic networks in a developmental atlas of maize. Science **353**:814–818.

Wang, X., Wang, H., Wang, J., Sun, R., Wu, J., Liu, S., Bai, Y., Mun, J.H., Bancroft, I., Cheng, F., et al. (2011a). The genome of the mesopolyploid crop species *Brassica rapa*. Nat. Genet. **43**:1035–1039.

Wang, Y., Wang, X., Tang, H., Tan, X., Ficklin, S.P., Feltus, F.A., and Paterson, A.H. (2011b). Modes of gene duplication contribute differently to genetic novelty and redundancy, but show parallels across divergent angiosperms. PLoS One **6**:e28150.

Wang, Z., Hobson, N., Galindo, L., Zhu, S., Shi, D., McDill, J., Yang, L., Hawkins, S., Neutelings, G., Datla, R., et al. (2012). The genome of flax (*Linum usitatissimum*) assembled *de novo* from short shotgun sequence reads. Plant J. **72**:461–473.

Wang, W., Haberer, G., Gundlach, H., Gläßer, C., Nussbaumer, T., Luo, M.C., Lomsadze, A., Borodovsky, M., Kerstetter, R.A., Shanklin, J., et al. (2014). The *Spirodela polyrhiza* genome reveals insights into its neotenous reduction fast growth and aquatic lifestyle. Nat. Commun. **5**:3311.

Wang, Y., Yu, H., Tian, C., Sajjad, M., Gao, C., Tong, Y., Wang, X., and Jiao, Y. (2017). Transcriptome association identifies regulators of wheat spike architecture. Plant Physiol. **175**:746–757.

Waters, A.J., Makarevitch, I., Noshay, J., Burghardt, L.T., Hirsch, C.N., Hirsch, C.D., and Springer, N.M. (2017). Natural variation for gene expression responses to abiotic stress in maize. Plant J. **89**:706–717.

Wendel, J.F., Jackson, S.A., Meyers, B.C., and Wing, R.A. (2016). Evolution of plant genome architecture. Genome Biol. **17**:37.

Woodhouse, M.R., Schnable, J.C., Pedersen, B.S., Lyons, E., Lisch, D., Subramaniam, S., and Freeling, M. (2010). Following tetraploidy in maize, a short deletion mechanism removed genes preferentially from one of the two homeologs. PLoS Biol. **8**:e1000409.

Woodhouse, M.R., Cheng, F., Pires, J.C., Lisch, D., Freeling, M., and Wang, X. (2014). Origin, inheritance, and gene regulatory consequences of genome dominance in polyploids. Proc. Natl. Acad. Sci. USA **111**:5283–5288.

Yang, J., Liu, D., Wang, X., Ji, C., Cheng, F., Liu, B., Hu, Z., Chen, S., Pental, D., Ju, Y., et al. (2016). The genome sequence of allopolyploid *Brassica juncea* and analysis of differential homeolog gene expression influencing selection. Nat. Genet. **48**:1225–1232.

Yim, W.C., Lee, B.M., and Jang, C.S. (2009). Expression diversity and evolutionary dynamics of rice duplicate genes. Mol. Genet. Genomics **281**:483–493.

Zhang, G.Q., Liu, K.W., Li, Z., Lohaus, R., Hsiao, Y.Y., Niu, S.C., Wang, J.Y., Lin, Y.C., Xu, Q., Chen, L.J., et al. (2017a). The *Apostasia* genome and the evolution of orchids. Nature **549**:379–383.

Zhang, Y., Ngu, D.W., Carvalho, D., Liang, Z., Qiu, Y., Roston, R., and Schnable, J. (2017b). Differentially regulated orthologs in sorghum and the subgenomes of maize. Plant Cell **29**:1938–1951.

Zheng, D., Ye, W., Song, Q., Han, F., Zhang, T., and Chen, Z.J. (2016). Histone modifications define expression bias of homeologous genomes in allotetraploid cotton. Plant Physiol. **172**:1760–1771.

Zimin, A.V., Puiu, D., Hall, R., Kingan, S., Clavijo, B.J., and Salzberg, S.L. (2017). The first near-complete assembly of the hexaploid bread wheat genome, *Triticum aestivum*. Gigascience **6**:1–7.