Evolution of Gene Duplication in Plants[1][OPEN]

Nicholas Panchy, Melissa Lehti-Shiu, and Shin-Han Shiu*

Genetics Program (N.P., S.-H.S.) and Department of Plant Biology (M.L.-S., S.-H.S.), Michigan State University, East Lansing, Michigan 48824

ORCID IDs: 0000-0002-1551-3517 (N.P.); 0000-0003-1985-2687 (M.L.-S.); 0000-0001-6470-235X (S.-H.S.).

Ancient duplication events and a high rate of retention of extant pairs of duplicate genes have contributed to an abundance of duplicate genes in plant genomes. These duplicates have contributed to the evolution of novel functions, such as the production of floral structures, induction of disease resistance, and adaptation to stress. Additionally, recent whole-genome duplications that have occurred in the lineages of several domesticated crop species, including wheat (*Triticum aestivum*), cotton (*Gossypium hirsutum*), and soybean (*Glycine max*), have contributed to important agronomic traits, such as grain quality, fruit shape, and flowering time. Therefore, understanding the mechanisms and impacts of gene duplication will be important to future studies of plants in general and of agronomically important crops in particular. In this review, we survey the current knowledge about gene duplication, including gene duplication mechanisms, the potential fates of duplicate genes, models explaining duplicate gene retention, the properties that distinguish duplicate from singleton genes, and the evolutionary impact of gene duplication.

Distinct from other eukaryotic genomes, plant genomes tend to evolve at higher rates, leading to higher genome diversity (Kejnovsky et al., 2009; Murat et al., 2012). For example, differences in genome size between closely related plant species are much larger than between other closely related eukaryotes. Among dicotyledonous species that diverged approximately 150 million years ago (MYA), genome size ranges from merely 63 Mb in the carnivorous *Genlisea margaretae* (Greilhuber et al., 2006) to approximately 150 Gb in the canopy plant *Paris japonica* (Pellicer et al., 2010). This 2,000-fold difference in genome size among dicots is in stark contrast to that observed among the mammalian species that also radiated approximately 150 MYA (Warren et al., 2008), where genome size ranges from approximately 1.6 Gb in Carriker's round-eared bat (Smith et al., 2013) to approximately 8 Gb in the tetraploid red viscacha rat (Gallardo et al., 1999).

Plant genomes also have an abundance of duplicate genes. Whole-genome duplication (WGD) has occurred multiple times over the past 200 million years of angiosperm evolution (Lyons et al., 2008; Soltis et al., 2009, 2014; Lee et al., 2013; Renny-Byfield and Wendel, 2014), and genomic sequencing continues to reveal new events (Velasco et al., 2010; D'Hont et al., 2012; Wang et al., 2012; Lu et al., 2013; Myburg et al., 2014; Wang et al., 2014b). In contrast, the most recent WGD event occurred approximately 450 MYA in the lineage leading to humans (Panopoulou et al., 2003; Dehal and Boore, 2005) and approximately 200 MYA in the budding yeast lineage (Wolfe and Shields, 1997; Kellis et al., 2004). Strikingly, many plant species also comprise mixed populations of diploid and polyploid individuals, illustrating the prevalence of polyploidy in plants (Husband et al., 2013). For example, 2.4% of *Lythrum salicaria* populations have both diploid and polyploid individuals (Kubatova et al., 2008), and this percentage is even higher (greater than 60%) for *Chamerion angustifolium* (Sabara et al., 2013) and *Actinidia chinensis* (Li et al., 2010).

WGD, or polyploidization, is an extreme mechanism of gene duplication that leads to a sudden increase in

ADVANCES

- On average, 65% of annotated genes in plant genomes have a duplicate copy. Of these, most were derived from WGD, consistent with the prevalence of paleopolyploidization events in the land plant lineage.
- Multiple mechanisms contribute to duplicate retention. Notably, mechanisms that do not require the evolution of new functions (e.g. dosage balance) may play an important role in the initial retention of duplicate genes.
- Retained duplicates and those that have reverted back to single copy can be predicted with models that incorporate gene functions and multiple sequence properties.
- Fitness differences between two locally adapted plant populations can be explained by genetic variation in duplicate genes involved in abiotic stress tolerance.

[1] This work was supported by the National Science Foundation (grant nos. MCB–1119778 and IOS–1126998 to S.-H.S.).

* Address correspondence to shius@msu.edu.

N.P. and S.-H.S. analyzed the data; N.P., M.L.-S., and S.-H.S. wrote the article.

[OPEN] Articles can be viewed without a subscription.

www.plantphysiol.org/cgi/doi/10.1104/pp.16.00523

both genome size and the entire gene set. However, it is not the only mechanism that gives rise to duplicated genes. In general, gene duplication generates two gene copies; this theoretically allows one or both to evolve under reduced selective constraint and, on some occasions, to acquire novel gene functions that contribute to adaptation. There is little question that duplicate genes have contributed to novel traits over the course of plant evolution (Van de Peer et al., 2009b). Through comparative analyses of an ever-increasing number of plant genome sequences and functional genomic data sets, we now have an unprecedented understanding of how genes are duplicated, how duplicated genes evolve new functions, and the impact of gene duplication on genome evolution (Conant and Wolfe, 2008; Freeling et al., 2015; Soltis et al., 2015).

Gene duplication is but one type of genomic change that can lead to evolutionary novelties. Novel functions can arise from the co-option of existing genes (True and Carroll, 2002), new genes can arise de novo from intergenic space (Tautz and Domazet-Lošo, 2011; Schlötterer, 2015), and new transcriptional regulatory sites can come into existence that alter gene expression (Wray et al., 2003). In addition, although in this review we focus only on genes, the duplication of other genomic features, including regulatory regions (Nourmohammad and Lässig, 2011), transposable elements (TEs; Lisch, 2013), and repeat elements (Sharopova, 2008), has been reported to influence gene expression and function. Nonetheless, gene duplication remains of specific interest both because of the abundance of plant gene duplicates and their potential to contribute to plant novelties. The goal of this review is to provide an overview of our current state of knowledge about plant gene duplication and its significance. We first focus on the prevalence of gene duplication in plants and the mechanisms that contribute to gene duplication. We then discuss the fate of duplicate genes and the factors that influence whether a duplicate is retained or not. Finally, we consider the influence of duplicate genes on the evolution of plant species and agronomically important traits.

PREVALENCE AND MECHANISMS OF DUPLICATION

Predominance of Duplicate Genes in Plant Genomes

In the green lineage, gene numbers range from 8,166 in the unicellular green alga *Ostreococcus tauri* (Derelle et al., 2006) to approximately 95,000 in bread wheat (*Triticum aestivum*; Brenchley et al., 2012). What proportion of genes within each species have shared common ancestry due to duplication (Fitch, 1970)? Although a paralog is well defined conceptually, the criteria (e.g. the threshold sequence similarity) and data sets used to identify paralogs vary among studies, and thus direct comparisons are difficult. For example, between 16% (barley [*Hordeum vulgare*]) and 49% (rice [*Oryza sativa*]) of plant genes were defined as paralogous based on transcript data (Blanc and Wolfe, 2004a); however, genome sequence data have yielded paralog frequencies as high as approximately 75% in soybean (*Glycine max*; Schmutz et al., 2010). In Arabidopsis (*Arabidopsis thaliana*), the estimate of duplicate gene content ranges from 47% (Blanc and Wolfe, 2004a) to 63% (Ambrosino et al., 2016) due to differences in gene models, methodology, and parameters (i.e. similarity cutoffs).

To obtain a comparable estimate of duplicate gene number across plant genomes, we applied a common methodology and similarity threshold to identify duplicate genes in 41 sequenced land plant genomes (Fig. 1). On average 64.5% of plant genes are paralogous, ranging from 45.5% in the bryophyte *Physcomitrella patens* to 84.4% in apple (*Malus domestica*). Given that ancient and/or fast-evolving paralogs are not easily detected due to sequence divergence, these percentages are likely underestimates. Total genic content is correlated significantly with both paralog content ($r^2 = 0.46$, $P < 7e\text{-}6$) and the presence of a reported polyploidization event ($r^2 = 0.35$, $P < 8e\text{-}4$), demonstrating the large contribution of duplication, particularly WGD, to differences in gene content among plant species.

Another way to illustrate the preponderance of plant duplicates is to look at the number of paralogs within gene families (Dayhoff, 1976). In a survey of eight diverse plant species, the percentage of genes belonging to gene families ranges from 40% in the green alga *Chlamydomonas reinhardtii* to 95% in the lycophyte *Selaginella moellendorffii*, with most species having in excess of 65% familial genes (Guo, 2013). Although the proportion of familial genes in plant genomes is high, there can be dramatic differences in the size of gene families across species due to lineage-specific expansions (Lespinet et al., 2002). For example, one of the largest families in plants is the protein kinase superfamily, which has 426 members in the unicellular green alga *C. reinhardtii* and 2,532 in *Eucalyptus grandis* (Lehti-Shiu and Shiu, 2012). Another example illustrating the variation in plant gene family size is the large difference in the number of transcription factors, which can differ more than 10-fold among plant species (Jin et al., 2014).

At the other extreme, some genes have few or no paralogs. For example, there is only one Arabidopsis gene encoding DNA gyrase A, despite the fact that repeated rounds of WGD would have generated gyrase A duplicates in the past. Thus, not all gene families are created equal. What contributes to these large differences in gene family size? Integrated analysis of gene family and functional annotation data led to the finding that plant genes involved in transcriptional regulation, signal transduction, and stress response tend to have paralogs (Blanc and Wolfe, 2004b; Maere et al., 2005; Shiu et al., 2005; Hanada et al., 2008) but those involved in essential functions, such as genome repair, genome duplication, and organelles, tend not to (Li et al., 2016). The correlation between duplication and function also appears to be influenced by how duplicates are made (Hanada et al., 2008). For example, transcription factors

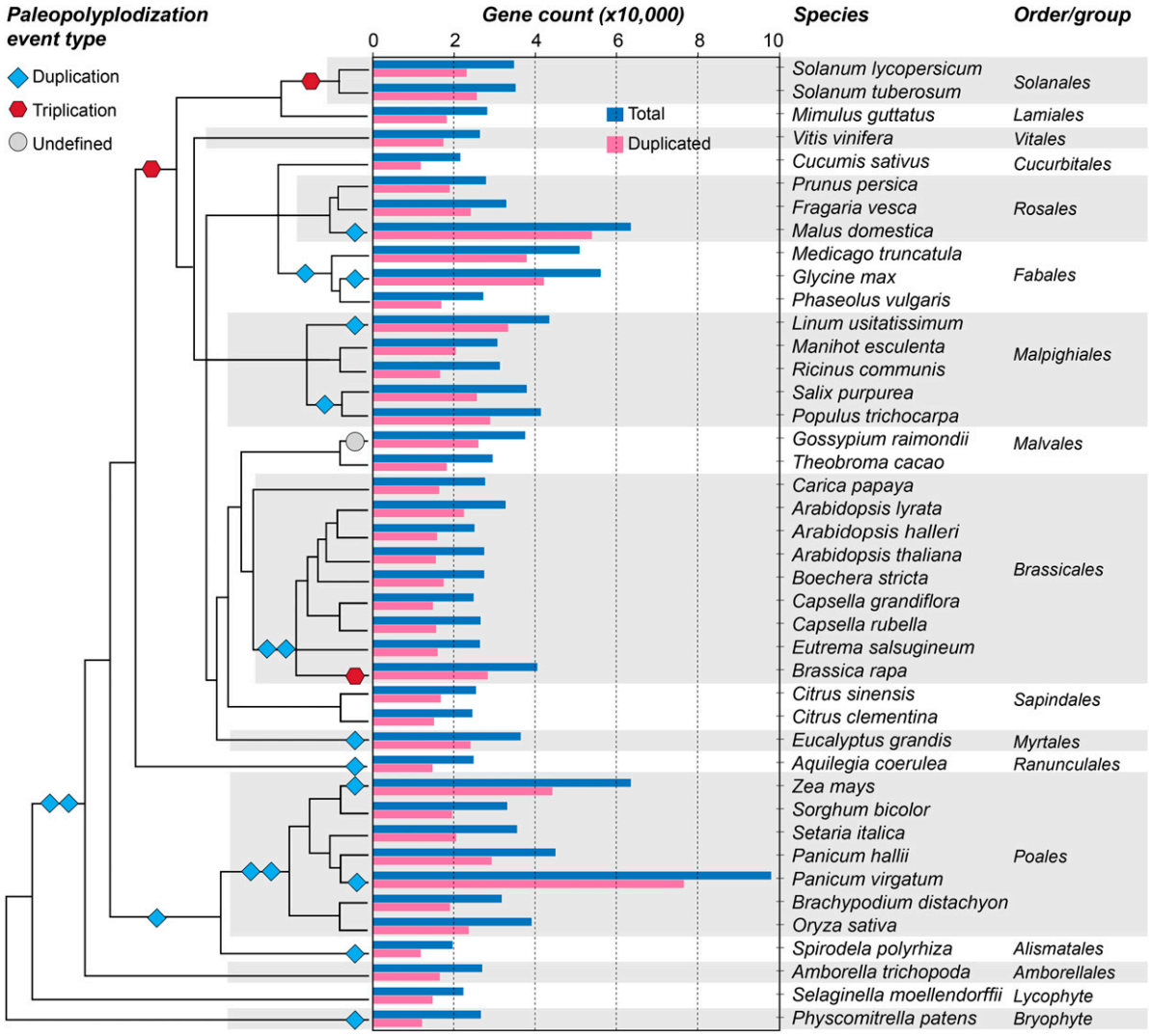

Figure 1. Duplication events and paralagous gene content in selected plant species. Left, Phylogeny of selected plant species. Duplication (squares), triplication (hexagons), and undefined (circles) polyploidization events are indicated on the tree. Middle, Total (blue) and duplicated (pink) gene numbers in each species. A gene is regarded as duplicated if it is significantly similar to another gene in a BLAST (Altschul et al., 1997) search (identity $\geq 30\%$, aligned region ≥ 150 amino acids, expect value $\leq 10^{-5}$). Right, Species names. The data used to generate this figure were obtained from CoGe and Phytozome 11 (Lyons and Freeling, 2008; Lyons et al., 2008; Goodstein et al., 2012) as well as genome annotations for *Eucalyptus grandis* (Myburg et al., 2014), *Panicum virgatum* (Lu et al., 2013), *P. patens* (Rensing et al., 2008), *Salix purpurea* (Phytozome), *Populus trichocarpa* (Tuskan et al., 2006), and *Spirodela polyrhiza* (Wang et al., 2014b). References for the information aggregated on CoGe can be found at https://genomevolution.org/wiki/index.php/Plant_paleopolyploidy.

have higher than average retention after WGD (Maere et al., 2005) but not after local duplication (Hanada et al., 2008). Therefore, to understand how duplication impacts gene content, it is necessary to know how genes are duplicated.

Mechanisms of Gene Duplication

Duplication is a form of mutation in which a genomic region is replicated and, in some cases, is inserted into a physically separate location. Multiple mechanisms contribute to gene duplication (Fig. 2). Considering the impact on gene content, the most dramatic form of gene duplication involves duplication of an entire chromosome or the whole genome (Fig. 2A). Ancient WGD events have taken place in the common ancestors of seed plants (approximately 340 MYA) and of angiosperms (approximately 170 MYA; Jiao et al., 2011). Subsequently, three rounds of WGD events (referred to as α, β, and γ) took place in the Arabidopsis lineage (Blanc et al., 2003; Bowers et al., 2003). In some cases, WGD events involve genome triplication, as is the case for *Brassica rapa* (Lysak et al., 2005), the wild radish

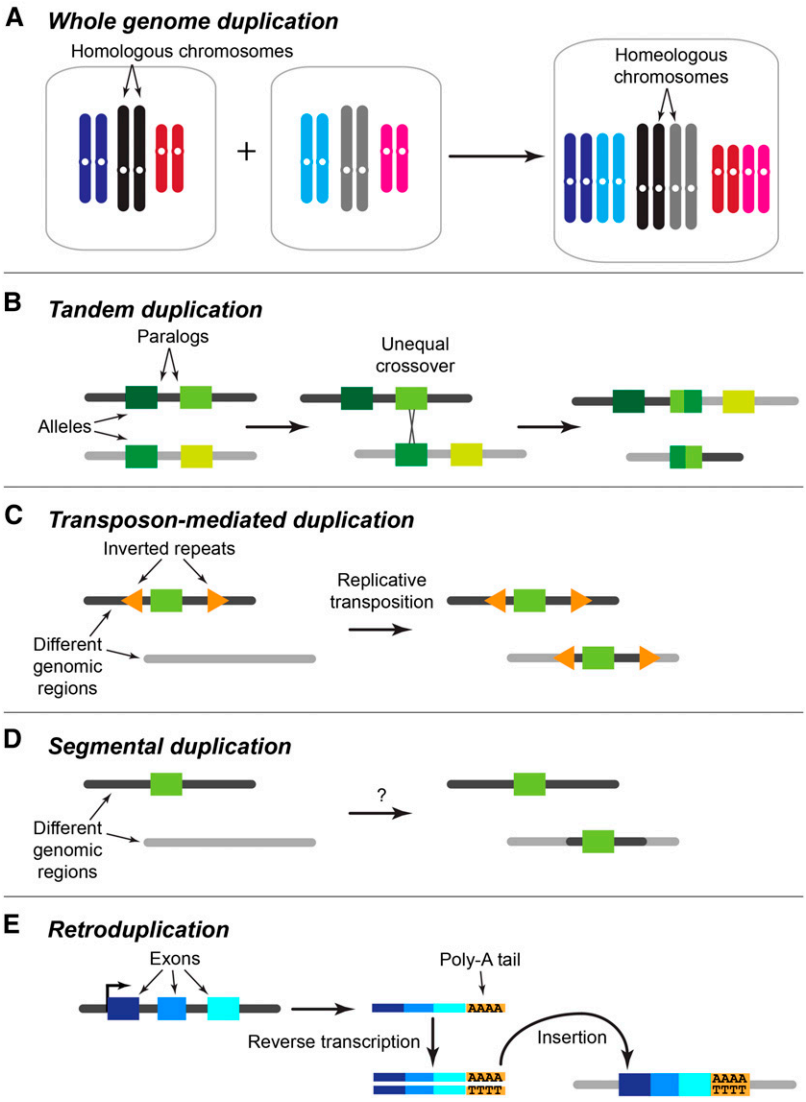

Figure 2. Mechanisms of gene duplication. A, WGD, or duplication of genes via an increase in ploidy. B, Tandem duplication, or duplication of a gene via unequal crossing over between similar alleles. C, Transposon-mediated duplication, or duplication of a gene associated with a TE via replicative transposition. D, Segmental duplication, or duplication of genes via replicative transposition of LINE elements in the human genome. The extent and causative elements of segmental duplication are ill defined in plants. E, Retroduplication, or duplication of a gene via reverse transcription of processed mRNA.

Raphanus raphanistrum (Moghe et al., 2014), and bread wheat (Salse et al., 2008). Even higher levels of polyploidization also have been observed (e.g. octoploid cultivated strawberry [*Fragaria* × *ananassa*]; Byrne and Jelenkovic, 1976). Paleopolyploidy events are thought to be rare, as only a handful of recognizable events occurred over the past 200 MYA (Van de Peer et al., 2009a). Nonetheless, the frequency of paleopolyploidization is much higher in plants than in any other eukaryotic lineage (Otto and Whitton, 2000). In addition, although the survival rate of nascent polyploids remains largely unclear (Soltis et al., 2015), diploid and polyploid cytotypes frequently coexist (Ramsey and Schemske, 1998). Given the frequency of ancient and more recent WGD events in plants, it is not surprising that WGD accounts for the majority of duplicate genes. In Arabidopsis, for example, approximately 60% of genes have at least one paralog in a corresponding syntenic block derived from one of three WGD events (Bowers et al., 2003).

The contribution of WGD to existing duplicates is likely much higher than estimated based on analyses of syntenic duplicates, because some WGD syntenic blocks, particularly those arising from older events, are no longer recognizable due to genome rearrangements, insertions, and deletions. Thus, some of the duplicates that cannot currently be ascribed to WGD may actually be WGD duplicates. Alternatively, they could be derived from subgenomic duplication events, which include tandem duplication (Zhang, 2003), duplication mediated by TEs (Jiang et al., 2004), segmental duplication (Bailey et al., 2002), and retroduplication (Drouin and Dover, 1990; Brosius, 1991). Tandem (or local) duplication (Fig. 2B) results from unequal crossing-over events and leads to a cluster of two to many paralogous sequences with no or few intervening gene sequences (Zhang, 2003). The number of tandem duplicates in plants varies widely, from 451 (4.6% of gene content) in *Craspedia variabilis* to 16,602 (26.1%) in apple (Yu et al., 2015). In Arabidopsis, the proportion of tandem

duplicates is close to the average of approximately 9% observed for 39 plant species. One challenge in estimating the contribution of tandem duplication to duplicate gene content is that, in the case of recent duplication events, paralogous copies may be misannotated as a single gene. For example, there are two *SEC10* paralogous genes involved in exocytotic vesicle fusion, but these genes make up only one locus in the assembled Arabidopsis genome (Vukašinović et al., 2014). Thus, misassembly contributes to an underestimate of tandem genes, and this issue is likely significant as most plant genomes are of draft quality.

In contrast to tandem duplication, which takes place locally, other subgenomic duplication mechanisms form dispersed duplicates. These mechanisms likely involve repetitive sequences and/or replicative transposition by TEs (Alleman and Freeling, 1986; Kapitonov and Jurka, 2007). In the case of TE-mediated gene duplication (Fig. 2C) in plants, gene capture by Mutator-like elements (MULEs) is the most prominent example (Bennetzen, 2005). There are approximately 3,000 rice Pack-MULEs that collectively contain approximately 1,000 fragmented or whole-gene sequences (Jiang et al., 2004). There are a similar number of Helitrons (approximately 2,800) in maize (*Zea mays*; Du et al., 2009) but only 46 Pack-MULEs in Arabidopsis (Jiang et al., 2011). This difference potentially reflects the historical difference in TE activity. Interestingly, Pack-MULEs preferentially obtain genic sequences (Ferguson et al., 2013), and a subset of Pack-MULE-carried genes are expressed and appear to be under selection (Hanada et al., 2009c). However, the mechanisms underlying this preferential acquisition and the functions of the acquired genes remain unclear. Similarly, in mammals, long interspersed nuclear elements and long terminal repeat retroposons are implicated in the generation of recent duplicates (Bailey et al., 2003; She et al., 2008) in a process referred to as segmental duplication (Fig. 2D; Bailey et al., 2002). Note that this is distinct from the original use of segmental duplication in the plant literature, which referred to rearranged genomic regions derived from WGDs (Arabidopsis Genome Initiative, 2000). Although both Pack-MULEs and mammalian segmental duplicates are associated with TEs, it remains unclear if they are generated via similar mechanisms.

Another TE-associated mechanism that generates dispersed duplicates is retroduplication (Fig. 2E; Drouin and Dover, 1990; Brosius, 1991). Here, mRNAs are reverse transcribed into DNA and inserted into the genome. These duplicate genes are referred to as retrogenes. Similar to Pack-MULEs, more retrogenes are found in rice (1,235) than in Arabidopsis (251; Wang et al., 2006; Abdelsamad and Pecinka, 2014), which also may reflect differences in past and current TE activity. In *Drosophila melanogaster*, retrogenes tend to be derived from genes that are highly expressed in germ-line tissues (Langille and Clark, 2007). However, because the regulatory sequences in the promoter are usually not duplicated, retrogenes were initially considered dead on arrival (Graur et al., 1989). However, there are several examples of functional retrogenes in mammals and *D. melanogaster* (Kaessmann et al., 2009). There is also evidence that plant retrogenes may be functional. Rice retrogenes have persisted longer than would be expected for nonfunctional elements and are under selection (Wang et al., 2006). In addition, studies in rice and Arabidopsis found that between one-fourth and one-third of retrogenes, respectively, have similar expression patterns to their parental genes (Sakai et al., 2011; Abdelsamad and Pecinka, 2014), and in Arabidopsis specifically, retrogene expression is up-regulated in pollen (Abdelsamad and Pecinka, 2014). However, while the functions of a few retrogenes have been examined in Arabidopsis (Abdelsamad and Pecinka, 2014), the overall functional significance of these duplicates has yet to be fully demonstrated.

Taken together, WGDs and tandem duplications account for the majority of plant duplicates, but TE-based mechanisms and retroduplication also generate a significant number of duplicates. It should be noted that, for example, in Arabidopsis, these mechanisms combined account for approximately 70% of duplicate genes. It remains to be determined if the remaining 30% were generated by some unknown mechanism or if there was a failure in assigning a duplication mechanism either due to the age-related erasure of specific signatures (i.e. synteny, proximity, and repeats) or too-stringent methods used to assign a mechanism.

GENE DUPLICATE LONGEVITY AND PSEUDOGENIZATION

Half-Life of Duplicate Genes

Despite the contribution of multiple duplication mechanisms and the variance in genome size, plant gene content remains relatively similar across land plant species. Considering that at least five rounds of WGD took place in the land plant lineage leading to Arabidopsis (Bowers et al., 2003; Jiao et al., 2011) and assuming that the common ancestor of land plants had approximately 10,000 genes, the number of genes in extant species would be 320,000 even without taking into account other duplication mechanisms. This expected gene number is approximately 10 times higher than the actual gene number in Arabidopsis and indicates extensive gene loss over time. Thus, although some duplicates have survived over millions to hundreds of millions of years, the predominant fate of most duplicates is loss (Li, 1983; Maere et al., 2005; Hanada et al., 2008). The preponderance of duplicates in plant genomes is driven mainly by the high rate of duplications over evolutionary time accompanied by the preferential retention of some duplicates.

How long will a duplicate survive after duplication? Assuming that the mutational process that leads to duplicate loss is stochastic, the longevity of a duplicate gene can be estimated in the form of half-life (i.e. the amount of time for half of the duplicates derived from a single event [e.g. WGD] to be lost; Lynch and Conery,

2000). The genome-wide half-life of Arabidopsis duplicates is estimated to be 17.3 million years (Lynch and Conery, 2003). For example, if a WGD event happened in the Arabidopsis lineage 17.3 MYA, we would expect that approximately 50% of the duplicates from that event will have been lost. As mentioned earlier, there have been multiple WGDs in the Arabidopsis lineage, so we can evaluate the consistency of duplicate half-life by considering these events independently. The most recent α WGD took place approximately 50 to 65 MYA (Bowers et al., 2003; Beilstein et al., 2010), and the observed duplicate survival rate ranges from 13.3% (Blanc et al., 2003) to 16.3% (Maere et al., 2005). Based on this information, the half-life estimate of α WGD duplicates is 17.2 to 24.8 million years. Although this is not far off from the genome-wide estimate of 17.3 million years (Lynch and Conery, 2003), α duplicates in general have a longer half-life than when all duplicates are considered. How about a more ancient WGD? The γ duplication likely took place approximately 140 MYA (Bowers et al., 2003; Moore et al., 2007), and 4.4% of duplicates are still retained (Maere et al., 2005). This implies a longer half-life, 31.3 million years, compared with α duplicates. On the other hand, certain gene families across angiosperms show bias against having paralogs from older WGD events, suggesting that their half-lives are shorter than average (Li et al., 2016). This difference in estimated half-life between the α and γ WGD events suggests that duplicate longevity is not constant over time; the rate of duplicate loss appears to decrease as the time since duplication increases, perhaps because a greater proportion of older duplicates are retained due to selective constraints.

Based on studies comparing duplicate and singleton (no closely related paralog) genes, the longevity of duplicates may be influenced by molecular and biological functions (Maere et al., 2005; Hanada et al., 2008), structural features (Jiang et al., 2013), number of protein interactions (Makino and McLysaght, 2012), and parent of origin (Song et al., 1995). Additionally, duplication mechanisms have different impacts on gene content. For example, WGD increases gene content dramatically but happens relatively infrequently. In contrast, tandem duplication, although affecting a limited number of genes, can increase or decrease gene number every meiotic division. These types of features that may affect duplicate retention are not taken into account when estimating genome-wide half-life (Lynch and Conery, 2003), and thus significant deviations are expected. Nonetheless, whether certain types of duplicates have longer or shorter half-lives, even the most conservative estimates suggest that most duplicates are lost relatively quickly after they are generated.

Mechanisms of Duplicate Gene Loss

The process of duplicate loss may involve deletion of the entire duplicate sequence and/or pseudogenization through loss-of-function mutations (Fig. 3A). If two duplicate genes are completely identical, there should not be a penalty for deleting either copy. In reality, however, duplicates are rarely equal. For example, copies derived from TE-mediated duplication, retroduplication, or tandem duplication may be missing parts of the parent gene-coding and/or regulatory regions. In these cases, loss of the new duplicate will likely incur no fitness penalty. Even in WGD, particularly when it involves allopolyploidy (merging of two related, but not identical, genomes), the patterns of duplicate loss are far from random (Thomas et al., 2006). The wholesale loss of duplicates is an important feature of fractionation, the reduction of genic content post WGD (Freeling et al., 2012). Analysis of syntenic blocks produced by the α WGD in Arabidopsis revealed that genes in one duplicate block were lost preferentially (Thomas et al., 2006). This fractionation bias has been observed in several plant species, including *Gossypium raimondii* (diploid cotton; Renny-Byfield et al., 2015), maize (Schnable et al., 2011), and *B. rapa* (Cheng et al., 2012). Importantly, this bias applies only to gaps in syntenic blocks that span multiple genes but not to gaps resulting from deletions of single genes (Thomas et al., 2006). Simulations of fractionation based on yeast data suggest that the observed bias in gene loss requires deletion events covering multiple genes and that random, single gene losses alone do not explain the observed pattern of fractionation (van Hoek and Hogeweg, 2007). What is the basis for fractionation bias? Looking across multiple WGD events, fractionation appears biased against duplicates with reduced expression level and promoter complexity (Schnable et al., 2012). Supporting this idea, analysis of lowly expressed genes in Arabidopsis suggests that they may be undergoing more rapid divergence and possibly pseuodgenization (Yang et al., 2011).

Nonfunctional duplicates are not always deleted; plant genomes are littered with thousands of apparently degenerated, nonfunctional duplicates referred to as pseudogenes (Benovoy and Drouin, 2006; Guo et al., 2009; Zou et al., 2009a). Pseudogenes are identified based on their similarity to annotated genes and the presence of disabling mutations (e.g. premature stop codons and frame shifts in protein-coding genes) that lead to presumed loss of function (Vanin, 1985). Although pseudogenes are presumably nonfunctional, a small subset of pseudogenes in rice and Arabidopsis are clearly expressed (Yamada et al., 2003; Thibaud-Nissen et al., 2009; Zou et al., 2009a). There are three potential explanations for pseudogene expression. First, some pseudogenes may have been falsely predicted due to misannotation. An example is the rice *ent-KAURENE SYNTHASE LIKE2* gene, which had mispredicted coding regions (Tezuka et al., 2015). This issue will likely be less prominent as transcript-based annotations improve (Law et al., 2015). The second explanation is that some pseudogenes may still be functional as truncated proteins or as RNA. For example, apomixis in the grass *Paspalum simplex* is hypothesized to be the consequence of antisense regulation by the transcript of a

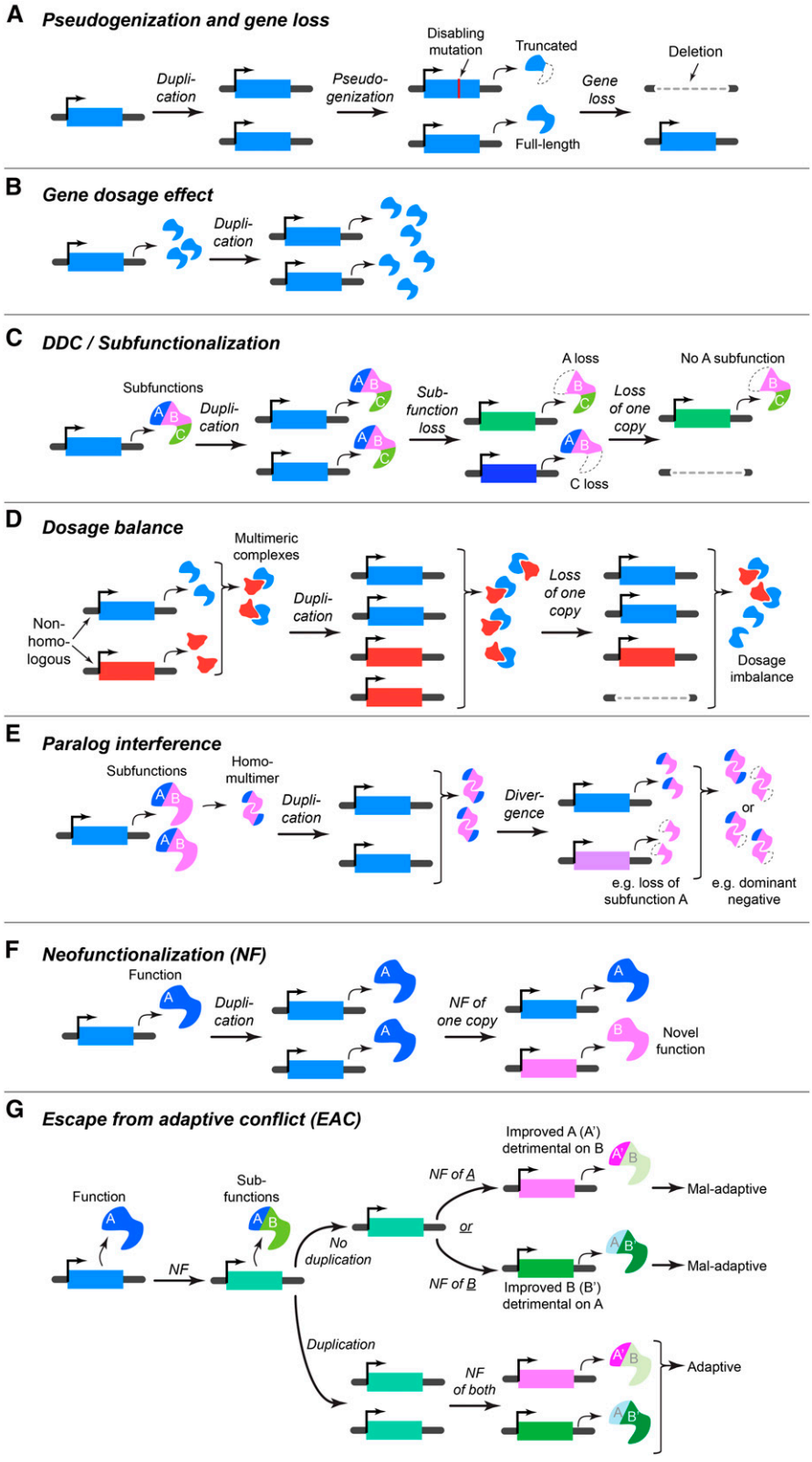

Figure 3. Potential fates of duplicate genes. Duplicate genes can be pseudogenized/lost (A), retained by selection on existing functions (B–E), or retained by selection on novel functions (F and G). Models of selection on existing function include the following: gene dosage, or retention of both duplicates because of a beneficial increase in expression (B); duplication-degeneration-complementation (DDC)/subfunctionalization, or retention of both duplicates to preserve the full complement of ancestral functions (C); dosage balance, or retention of both duplicates to maintain the stoichiometric balance (D); and paralog interference, or retention of both duplicates to prevent interference between the products of each paralog (E). Models of selection on novel functions include the following: neofunctionalization, or retention of both duplicates because of a gain of function post duplication (F); and escape from adaptive conflict (EAC), or retention of both duplicates that allows for the independent optimization of conflicting ancestral functions (G).

pseudogene related to the *ORIGIN RECOGNITION COMPLEX3* gene (Siena et al., 2016). However, there is no direct evidence that the pseudogene is responsible for the antisense transcript. Finally, some pseudogenes may have become pseudogenized relatively recently and are in the process of complete decay. This is

consistent with the finding that expressed pseudogenes tend to be derived from more recent duplication events (Zou et al., 2009a). In addition, pseudogenes, expressed or not, tend to have elevated nonsynonymous (amino acid-changing) substitution rates (Zou et al., 2009a), indicating that they are not subject to the same degree of selective pressure as their intact relatives.

Do duplicates have similar propensities for pseudogenization? In *B. rapa* and *R. raphanistrum*, which experienced a recent genome triplication event, there are significantly more pseudogenes than in the related species Arabidopsis and *Arabidopsis lyrata*, which did not experience a recent WGD (Moghe et al., 2014). Also, there is a significant, positive correlation between gene family size and the number of pseudogenes within families (Zou et al., 2009a). Thus, in general, the more members of a family that are duplicated, the more losses occur. However, this correlation is far from perfect (Zou et al., 2009a), indicating that other factors are important. One such factor is gene function. For example, Arabidopsis pseudogenes tend to have functional relatives playing roles in disease resistance, specialized (secondary) metabolism, cell wall modification, and protein degradation (Zou et al., 2009a), but transcription factor and receptor-like kinase families tend not to have pseudogenes (Hanada et al., 2008). In addition, pseudogenes tend to be derived from tandem duplicates (Hanada et al., 2008), although this may be due to the higher rate of tandem duplication compared with other duplication mechanisms. Taken together, duplicate longevity depends on functional role and duplication mechanism, which necessarily means that there is a significant bias in the kinds of duplicates that are retained.

MECHANISMS FOR RETENTION OF DUPLICATE GENES

Genetic Drift and Genetic Redundancy

Over the course of plant evolution, hundreds of thousands of new genes were created by duplication, and most of these duplicates were lost over time. Nonetheless, considering that more than half of the gene content in most plant species consists of duplicates, some duplicates have clearly escaped this fate. Why do some duplicates persist while others are lost? Models for duplicate gene retention in general have been reviewed elsewhere (Innan and Kondrashov, 2010; Maere and Van de Peer, 2010); here, we will focus on examples of duplicate retention in plants (Fig. 3, B–G). It is important to note that these models are not mutually exclusive. For example, selection on both duplicates to maintain dosage balance (Fig. 3D) contributes to increased duplicate longevity, which may allow time for the evolution of novel functions (Fig. 3F; Veitia et al., 2013). Here, we discuss each model independently to emphasize the distinct mechanisms that contribute to duplicate retention. First, we discuss the idea that both duplicates are retained without a significant change in function, either because insufficient time has passed for deleterious mutations to accumulate or because there is selection pressure to retain redundant functions.

Assuming that mutations accumulate randomly and that selection is not a factor, genetic drift will be the dominant factor influencing the frequency of mutant alleles (Kimura, 1968, 1983). In this case, a mutation that appears in a gene would take approximately four N_e generations to become fixed (where N_e is effective population size; Kimura and Ohta, 1969). In Arabidopsis, which has an estimated N_e of 250,000 (Cao et al., 2011) and a winter annual life cycle for most accessions (Michaels et al., 2004), the time to fixation is approximately 1 million years. When a pair of duplicates is considered, the situation is more complicated because either copy can be lost without affecting fitness. Assuming that the time for a recent WGD to sweep through the population is negligible, the time to fixation is a function of N_e, the fitness effect of the loss of both duplicate genes, and the mutation rate at the duplicated loci (Kimura and King, 1979). The average time to fixation is estimated to be between three and 20 N_e generations. In Arabidopsis, this translates to an average fixation time of between 0.75 and 5 million years. Thus, it is expected that some duplicate genes potentially survive for several million years due to genetic drift and not because their presence is beneficial. In this situation, some duplicates may be decaying functionally even though there is no apparent sign of pseudogenization (Lehti-Shiu et al., 2015). However, if drift were the only factor affecting retention and the expected time to deletion was approximately 1 million years, we would expect a much lower genome-wide duplicate half-life (mean lifetime approximately 1.44 × half-life). Thus, a substantial number of duplicates are most likely under selection (i.e. loss of function in either of the duplicates is expected to reduce the fitness of the individual).

Alternatively, duplicate retention might occur via selection for genetic redundancy (or genetic buffering), where the effects of a null mutation are ameliorated (or buffered) due to the presence of an intact, duplicate copy (Zhang, 2012). The prediction is that developmental or physiological phenotypes are only obvious when a gene and one or more of its relatives (paralogs) are mutated (Nowak et al., 1997). Consistent with the idea of genetic redundancy, phenotypic effects when one copy of a duplicated pair is disrupted are significantly smaller compared with those observed when a singleton gene is disrupted in Arabidopsis (Hanada et al., 2009a). However, claims of genetic redundancy between duplicates thus far are based on the absence of gross morphological, developmental, and/or behavioral phenotypes in highly controlled environments. Thus, relatively subtle phenotypic changes or conditional phenotypes resulting from mutations in one duplicate copy, which may have fitness consequences, may not be detected. Additionally, although redundancy may be beneficial in a way that is analogous to a fail-safe in an engineered system

(McAdams and Arkin, 1999; Kitano, 2004; Kondrashov, 2010), it remains to be shown how selection can act in anticipation of future need to favor redundancy. Although long-term conservation of redundant duplicates is feasible in simulated situations, it requires perfect equivalency in gene functions and in mutation rates between the two duplicates (Nowak et al., 1997), which is highly unlikely. After gene duplication, various degrees of functional redundancy are expected. But the redundant functions can be present not necessarily because they are useful but because there has been insufficient time for their loss. Due to the challenges in experimentally assessing the functional and fitness contributions of duplicates, the true extent of genetic redundancy and whether redundancy is the cause or consequence of duplicate retention remain unclear.

Selection on Existing Functions

Duplicates can be retained without acquiring new functions via one of the following four mechanisms: gene dosage increase (Ohno, 1970), DDC (Force et al., 1999), gene balance (Freeling and Thomas, 2006), and paralog interference (Baker et al., 2013). Ohno (1970) recognized that, in situations where increased gene dosage confers an advantage by meeting metabolic demands, the presence of duplicate copies is beneficial (Fig. 3B). Unlike redundancy, the robustness of duplicates is clearly selectable (McAdams and Arkin, 1999; Kitano, 2004; Kondrashov, 2010); although the molecular function of the duplicates may be unchanged, the effect of increased dosage is new. Models of budding yeast gene networks suggest that WGD likely contributed to increased flux in the glycolytic pathways, which confers a fitness benefit in high-Glc environments (van Hoek and Hogeweg, 2009). Similarly, in Arabidopsis, duplicate retention after WGD is associated with reactions with high metabolic flux (Bekaert et al., 2011), suggesting that increased gene dosage results in increased metabolic activity, which may be beneficial. The impact of duplicate gene dosage is emphasized in a recent review (Conant et al., 2014).

In DDC or subfunctionalization (Fig. 3C; Force et al., 1999), after duplication of a multifunctional gene, each copy randomly loses subfunctions of the original gene (degeneration), and because each duplicate loses different subfunctions, both copies have to be kept to maintain the original, ancestral functionality (complementation). An important point to emphasize is that duplicate retention through DDC does not require any new evolved functions, just partitioning of the old ones. Evidence suggests that DDC has occurred at the level of protein function (Aklilu et al., 2014; Aklilu and Culligan, 2016), at the level of gene expression (Duarte et al., 2006; Throude et al., 2009; Zou et al., 2009a; Ma et al., 2015), and at both levels simultaneously (Geuten et al., 2011). Under DDC, subfunctions are assumed to be independently mutable, and it is expected that the ancestral functions are partitioned symmetrically among duplicates. Interestingly, the pattern of functional partitioning between duplicates tends to be highly asymmetric, where one copy tends to have significantly more subfunctions than its paralog, compared with what would be expected randomly (Zou et al., 2009b). Thus, either subfunctions are not independently mutable and/or other factors affect the partitioning of subfunctions.

An alternative model for duplicate retention is the gene balance hypothesis (Birchler et al., 2005; Birchler and Veitia, 2007, 2010). Under one version of this model, after WGD, duplicate genes that are dosage sensitive tend to be retained (dosage balance; Fig. 3D; Thomas et al., 2006). The idea is that duplication of a gene whose product has a greater number of molecular interactions (e.g. protein-protein or protein-DNA interactions) will lead to a dosage imbalance if all its interactors remain single copy. The individual harboring the duplicate will then have reduced fitness due to this imbalance. A higher degree of imbalance is expected for gene products with a higher number of interactions. In this situation, the duplication of just one gene will likely have a deleterious effect. But when this highly connected gene is duplicated along with its interactors in a WGD event, its retention is favorable because its removal would lead to imbalance. Consistent with gene balance, the expression patterns of highly interconnected genes tend to be more highly correlated than random duplicates (Lemos et al., 2004). Furthermore, there is a greater correlation between the expression patterns of WGD duplicates than between tandem duplicates (Casneuf et al., 2006; Arabidopsis Interactome Mapping Consortium, 2011).

Gene balance also can result from mechanisms other than dosage balance. In situations where molecular interaction is important for gene function, degenerative mutations in one duplicate copy may interfere with the functions of its paralog (paralog interference; Fig. 3E). These degenerative mutations will be selected against, leading to retention of both duplicates (Bridgham et al., 2008; Baker et al., 2013). Paralog interference is distinct from dosage balance in two respects. First, paralog interference occurs at the level of protein interaction and is independent of changes in the amount of gene products. Second, it is relevant specifically to situations where the formation of homomultimers is important for the ancestral gene function. The sequestration of interactors by the degenerate copy effectively creates an imbalance but not in dosage. Given that homomultimerization is prominent in multiple protein families such as transcription factors (Amoutzias et al., 2008), paralog interference may have a significant contribution to duplicate retention post WGD (Kaltenegger and Ober, 2015). Note that paralog interference also is distinct from DDC, which requires degenerative mutations to explain retention.

Contribution of Selection on New Functions to Duplicate Retention

Duplicates, although originally sharing the same functions, may acquire new functions. If these functions are beneficial, selection will act to retain both duplicates. Two models that explain duplicate retention due to the acquisition of novel adaptive functions are neofunctionalization (Ohno, 1970) and EAC (Des Marais and Rausher, 2008). Under the neofunctionalization model, one duplicate retains the ancestral function while its paralog gains a novel function (Fig. 3F). If the novel function contributes to better fitness, selection should maintain both duplicates. Note that determining whether neofunctionalization has taken place requires knowledge of gene functions prior to duplication. Some examples where neofunctionalization after duplication has likely contributed to duplicate retention include MADS box transcription factors involved in the evolution of novel floral structures (He and Saedler, 2005; Hu et al., 2015b; Zhang et al., 2015), 4,5-dioxygenase and cytochrome P450 genes that contribute to pigment variation in Caryophyllales (Brockington et al., 2015), and the recruitment of duplicated primary metabolite genes into specialized metabolite pathways (Durbin et al., 2000). At the gene expression level, it is estimated that approximately 10% of Arabidopsis duplicate genes have gained a novel response to stress conditions (Zou et al., 2009b), although it remains to be determined whether these novel responses are adaptive or not.

Similar to DDC, EAC (Hittinger and Carroll, 2007; Des Marais and Rausher, 2008) predicts subfunctionalization followed by the specialization of duplicates, but in EAC, novel functions arise prior to duplication (Fig. 3G). Subsequent subfunctionalization allows the separate functions of the ancestral gene to evolve independently. Thus, the distinguishing characteristic of EAC is improvement of the original ancestral function in one duplicate and of the novel function in the other (Des Marais and Rausher, 2008). EAC also is similar to the classic neofunctionalization model in that duplication allows further adaptive changes to accumulate, but the neofunctionalization model postulates that, after duplication, only one copy acquires a novel function. Under both neofunctionalization and EAC, duplicates are retained due to their adaptive contribution: individuals with duplicates that have novel and/or optimized functions are expected to have higher fitness compared with individuals with single copies. In contrast, under DDC, the duplicates do not provide an adaptive edge compared with the single-copy gene. Because of the requirement that the novel function existed prior to duplication, EAC is likely particularly relevant to genes that are nonessential and not highly conserved but are highly liable to selection (Sikosek et al., 2012). Examples of EAC in plants include improved enzymatic activity on ancestral flavonoid substrates after duplication of dihydroflavonol-4-reductase leading to adaptive changes in anthocyanin synthesis (Des Marais and Rausher, 2008) and salicylic acid/benzoic acid/theobromine enzymes, where improved enzymatic activities likely are under positive selection (Huang et al., 2012).

The models that we have presented here as distinct may actually involve a combination of different mechanisms. For example, EAC invokes both the neofunctionalization and DDC models to explain duplicate retention. Similarly, under the subneofunctionalization model (He and Zhang, 2005), the expected outcome of duplication is first the partitioning of ancestral functions followed by neofunctionalization in both duplicates. Subneofunctionalization is distinct from EAC, however, in that the total number of unique functions of the duplicate pair is expected to exceed the number of original functions. This is consistent with the finding that the numbers of protein interactions among two duplicate genes are higher than those of randomly paired singletons (He and Zhang, 2005). Paralog interference does not require adaptive change as a mechanism for duplicate retention, but subsequent neofunctionalization may resolve interference in a way that further increases the adaptive value of both duplicates (Baker et al., 2013; Kaltenegger and Ober, 2015). Taken together, the retention of a duplicate may involve any individual or combination of the above mechanisms, and it is of interest to know the relative contributions of each. Although some studies have explicitly compared different models of retention (Yang et al., 2006), the major obstacle to assessing the mechanisms of duplicate retention remains the inference of ancestral function, which requires a number of assumptions and is hypothetical. Thus, assessing the relative contributions of each duplication mechanism, or combination of mechanisms, to duplicate retention remains a major challenge.

PROPERTIES OF RETAINED DUPLICATES

Evolutionary Rate of Duplicates

After gene duplication, the rate of evolution (sequence substitutions) should increase, at least initially, because the presence of two copies relaxes selection against previously deleterious mutations (Scannell and Wolfe, 2008). Consistent with this expectation, duplicates display relaxed purifying selection (Carretero-Paulet and Fares, 2012) as well as differences in sequence structure, such as length of the coding region and the size and distribution of indels (Wang et al., 2013a). This increase in evolutionary rate is not necessarily equal for both copies: the proportion of WGD duplicates with evidence of asymmetric evolution ranges from 21.2% in maize to 68.3% in *P. patens* (Carretero-Paulet and Fares, 2012). Additionally, after a recent genome triplication approximately 25 MYA, 13% to 19% of *B. rapa* and *R. raphanistrum* paralogs evolved asymmetrically (Moghe et al., 2014). This pattern

is expected under the neofunctionalization model, where one copy maintains the original function while mutations that contribute to new functions accumulate in the other copy. However, this pattern also can be explained by the gradual loss of function in one copy: once one degenerative mutation has occurred in one duplicate, the chance for a second mutation to occur in the same duplicate is expected to be higher. Thus, for any particular case of asymmetry, the underlying cause needs to be determined, and neofunctionalization cannot be assumed by default.

Asymmetry in evolutionary rate demonstrates differential evolution between duplicate copies, but what about relative to the previous, unduplicated copy? To assess whether duplication is associated with altered evolutionary rates, three different comparisons can be made. First, a duplicate pair can be compared with the putative ancestral gene. This is exceedingly difficult due to the challenges associated with inferring ancestral function. Second, a duplicate pair can be compared with a closely related singleton (that was duplicated in the past but whose paralogs were lost). In this case, the duplicates and singletons have common ancestry and, presumably, similar functions. An exemplary study of this type provides evidence for asymmetric sequence evolution of both WGD and small-scale duplicates in multiple plant species (Carretero-Paulet and Fares, 2012). Third, all duplicates (genes with paralogs) can be compared with all singletons (those with no obvious paralog) in order to assess the average trend. In Arabidopsis, the nonsynonymous (amino acid-changing) substitution rates of duplicates tend to be lower compared with those of singletons (Yang and Gaut, 2011). The apparently more constrained evolution among duplicate genes can be the consequence of gene balance and/or paralog interference, but there can be other confounding factors that complicate the comparison of duplicates and singletons. For example, the evolutionary rate differences between duplicates and singletons can reflect differences in gene functions (discussed below) and sequence features such as protein domains, which tend to be longer, more numerous, and more highly conserved in duplicates than in singletons (Chapman et al., 2006). Additional sequence features differ between singleton and duplicate genes across multiple plant species (Jiang et al., 2013), but it remains to be seen whether these features contribute to the difference in evolutionary rate or are themselves consequences of duplication.

Expression Patterns of Duplicates

In addition to sequence differences, duplicate genes can have divergent expression patterns. At the transcriptional level, approximately 70% of duplicate pairs in Arabidopsis have significant differences in transcript levels (Ganko et al., 2007). In *Gossypium* (cotton), the transcript levels of 99.4% of the duplicates derived from a WGD event that occurred 60 MYA have diverged (Renny-Byfield et al., 2014). Expression divergence at the protein level also has been documented (Hu et al., 2015a). This divergence in expression may be under selection due to subfunctionalization and/or neofunctionalization. However, this difference in expression could result from fractionation among WGD duplicates (Schnable et al., 2011) that may not be subject to selection, at least initially. If expression differences were purely neutral, paralogs from younger duplication events would be expected to have more similar expression profiles. Yet the correlation between duplicate expression similarity and the timing of duplication (using synonymous substitution rate as a proxy) is weak, and the direction of correlation is not consistent between WGD (positive) and tandem (negative) duplicates (Haberer et al., 2004; Ganko et al., 2007). This is also true in rice (Li et al., 2009) and poplar (*Populus* spp.; Rodgers-Melnick et al., 2012). Thus, expression divergence is not an entirely neutral process. In contrast, there is a highly significant negative correlation between nonsynonymous substitution rate and expression similarity between duplicates (Ganko et al., 2007). That is, the more divergent the protein sequences are, the more similar the expression profiles. Assuming that the duplicates that contribute to this pattern are indispensable, it would appear that duplicates will be retained if they have sufficiently large differences in either expression profiles or sequences and, in some cases, a combination of both. Consistent with this notion, younger duplicates that are essential (lethal when mutated) do, in fact, have more divergent expression profiles compared with older, essential duplicate genes (Lloyd et al., 2015). In this case, the large difference in expression patterns of young duplicates contributes to the lack of buffering effect if one duplicate is lost.

Expression divergence between duplicates can arise due to differences at various stages of expression regulation. First, differences in cis-regulatory elements between Arabidopsis duplicate genes explain, in part, why duplicates have different responses to stressful environments (Zou et al., 2009b). In *Gossypium hirsutum* (allotetraploid cotton), 40% of homologs derived from a recent WGD display transcriptional divergence that can be attributed to cis-regulatory divergence between the diploid progenitors (Chaudhary et al., 2009). This result also is consistent with a recent study that found divergence in DNaseI footprints in the promoters of recently duplicated genes, suggesting that they are regulated by different sets of transcription factors (Arsovski et al., 2015). The degree of regulatory interaction divergence is dependent on the duplication mechanism: most WGD-derived duplicates share some regulatory interactions, while non-WGD duplicates tend not to have overlapping regulators (Arsovski et al., 2015). Gene body methylation is found to impact transcription (Lister et al., 2008; Takuno and Gaut, 2012), and divergence in methylation pattern between duplicates is significantly, although rather weakly, correlated with expression divergence in Arabidopsis (Wang et al., 2014a), rice (Wang et al., 2013b), and cassava

(*Manihot esculenta*; Wang et al., 2015a). Beyond transcriptional regulation, duplicates with divergent microRNA-binding sites tend to have more divergent expression profiles (Wang and Adams, 2015), and duplicates also can differ significantly in alternative splicing. For example, in hexaploid bread wheat, 42% and 61% of alternative splicing events differ between homologous gene pairs in chromosome 3A-3B and 3A-3D comparisons, respectively (Akhunov et al., 2013). In Arabidopsis, 85% of α WGD and 89% of tandem duplicates have divergent alternatively spliced forms (Tack et al., 2014).

Much like evolutionary rate, the expression of duplicates as a whole can be significantly different from that of singletons (Holstege et al., 1998; Seoighe and Wolfe, 1999). In plants, duplicates tend to be consistently more highly expressed than singletons in multiple plant species (Jiang et al., 2013). Conversely, while breadth of expression also is higher in duplicates in general, the trend is not universal. The differences in chromatin accessibility between duplicates and singletons may provide an explanation for expression level differences: in Arabidopsis, the promoter regions of duplicate genes have nearly twice as many DNase I hypersensitive sites compared with singleton genes (Arsovski et al., 2015). Taken together, these studies highlight the molecular mechanisms that underlie expression divergence between duplicates and between duplicates and singletons. The challenge now is to distinguish between the expression differences that contribute to differences in duplicate functions and those with no significant impact.

Functions of Duplicates

Given that some, although not all (e.g. gene balance), models of duplicate retention involve selection on gain- or loss-of-function mutations, we expect to find evidence of functional difference in duplicate pairs. For the purposes of this review, we classify functions into two different categories: molecular function and biological process-based function. Molecular function is defined as the molecular activity of a gene product (e.g. protein-protein interaction). Analysis of Arabidopsis interactome data revealed that duplicates tend to have different interaction partners (Carretero-Paulet and Fares, 2012; Guo et al., 2013). Nearly half of all WGD duplicate pairs are completely diverged, with no shared protein-protein interactions (Guo et al., 2013). Younger duplicate pairs tend to have more shared interactions, but a similar portion of young (92.7%) and old (97.3%) duplicates have at least some divergence in protein-protein interactions. Greater discordance in protein-protein interactions tends to be correlated with decreased expression similarity and a greater number of distinct protein domains (Guo et al., 2013). Duplicate genes as a whole encode proteins with significantly more protein-protein interactions compared with singleton genes (Alvarez-Ponce and Fares, 2012). However, while genes derived from WGD have greater regulation connectivity, tandem and WGD duplicates show no significant differences in the number of associated protein-protein interactions (Carretero-Paulet and Fares, 2012).

The function of a gene also can be defined as the impact of its gene product on a biological process. There is a large body of experimental evidence demonstrating how biological process functions differ between two plant duplicates. For example, paralogous MADS domain transcription factors control different aspects of plant development, and this functional divergence is due to changes in both promoter and coding regions (McCarthy et al., 2015). To summarize how the biological processes involving duplicates differ from those involving singleton genes, one approach is to examine annotated gene functions. Importantly, the functional differences between duplicates and singletons are heavily influenced by the duplication mechanism. For example, Arabidopsis WGD duplicates involved in signal transduction and transcriptional regulation tend to be retained (Blanc and Wolfe, 2004b; Seoighe and Gehring, 2004; Maere et al., 2005; Shiu et al., 2005). On the other hand, Arabidopsis and rice genes in Gene Ontology (GO; Ashburner et al., 2000) categories relevant to response to environmental stress tend to be tandemly duplicated (Rizzon et al., 2006; Hanada et al., 2008), and Arabidopsis genes that are transcriptionally responsive to stress have experienced lineage-specific expansion due mainly to tandem duplication (Hanada et al., 2008). To assess if the functional biases in duplicate retention are consistent across species, the enrichment of GO terms in WGD and tandem duplicates was evaluated in six flowering plants (Jiang et al., 2013). The overrepresentation of GO terms was found to be anti-correlated between WGD and tandem duplication within each species. However, only a subset GO terms (transcriptional regulation, ribosomes/translation, amino acid synthesis, and kinase activity) had the same pattern of enrichment in the majority of species (Jiang et al., 2013). This is consistent with another study that found differences between the functions of WGD and tandem genes but little correlation between the overall pattern of GO category enrichment in duplicate genes across four plant species (Carretero-Paulet and Fares, 2012). Conversely, gene families that are common across sequenced angiosperms are biased toward single-copy status and retained as such in all species except those with a recent WGD (less than 50 MYA; Li et al., 2016). Although GO-based analysis has been informative, there can be significant issues with its use (Rhee et al., 2008), particularly because there are differences in the quality of functional annotation for different genomes. Further meta-analysis using compiled experimental data across species, which is routinely done in evo-devo studies of floral evolution, for example (Fornoni et al., 2016), may provide the required resolution and accuracy to assess functional differences between duplicates and singletons.

Taken together, duplicate and singleton genes have significantly different sequence properties, expression patterns, molecular functions, and biological roles. These properties can be used to construct quantitative models that can predict whether a duplicate will be retained or not. Among the properties that distinguish duplicates from singletons, which ones are more important? Based on duplicate genes from six flowering plant species, the number of unique protein domains (another metric for the number of functions), nucleotide diversity, and GC content were found to be the most important predictors (Jiang et al., 2013). In another study, where multiple sequence properties, measures of conservation, molecular functions, and biological network information were consolidated in a single quantitative model to predict retention (Moghe et al., 2014), it was shown that no single property could accurately distinguish between duplicates and singletons. However, models that used only GO terms, sequence, and conservation features could predict duplicates equally well as models including the full set of features (Moghe et al., 2014), highlighting the importance of these features in duplicate retention.

EVOLUTIONARY, ECOLOGICAL, AND AGRONOMIC IMPACTS OF GENE DUPLICATION

Contribution of Duplicate Genes to Evolutionary Novelty

A common theme among models of duplicate retention is that, for both copies of a gene to be maintained, differences in function, expression, or interaction occur in one if not both paralogs. In some cases, duplicates will acquire novel functions and contribute to evolutionary novelties. For certain duplicates, evidence for a novel function can be seen in the morphological phenotypes caused when they are knocked out, and the presence of such a phenotype is correlated with the extent of sequence and expression divergence (Hanada et al., 2009b). The impact of duplication on evolution has been discussed in depth (Van de Peer et al., 2009b). Here, we focus on plant examples and classify evolutionary novelties into three types: (1) novel molecular function (e.g. expression in a novel context or interaction with a new protein); (2) novel plant structure/function that results from a new molecular function (e.g. a new floral organ or a new cell type); and (3) novel adaptive traits that result from a novel structure/function (e.g. disease resistance). This distinction is important because the novel functions/structures in types 1 and 2 need not be adaptive and, thus, may be novel without being selected for. What is the evidence that gene duplication has contributed to these three types of novelties in plants?

Novel molecular functions acquired after gene duplication can easily be seen in the context of gene expression. For example, Arabidopsis duplicates have likely accumulated novel environmental responses (Zou et al., 2009b) and novel developmental regulatory patterns (Liu et al., 2011). Novel leaf gene expression patterns also are observed for some maize duplicate genes (Hughes et al., 2014). However, a novel expression pattern is not necessarily adaptive. In the case of novel environmental responses, some of these responses may be adaptive (Zou et al., 2009b), but direct evidence is not available. Duplication also has contributed to the generation of novel metabolic functions. Specialized metabolism genes are retained preferentially following tandem duplication (Yu et al., 2015), and tandemly duplicated clusters have been found in multiple plant species (Kliebenstein and Osbourn, 2012; Nützmann and Osbourn, 2014). Interestingly, some specialized metabolic genes likely arose from neofunctionalized duplicates of primary metabolism genes (Qi et al., 2006), and further duplications of specialized metabolic genes have likely contributed to additional novel biochemical activities (Field and Osbourn, 2008; Takos et al., 2011). In some cases, these duplicates are implicated in defense against herbivores and microbes as well as in attracting pollinators (Mizutani and Ohta, 2010; Moghe and Last, 2015). However, the adaptive significance of most novel biochemical activities has yet to be demonstrated.

There also is evidence that duplicate genes have contributed to novel plant structures/functions, and in some cases, there is clear adaptive significance. A key example is the specification of floral organ identity by MADS box transcription factors (Bowman et al., 1989; Sommer et al., 1990; Coen and Meyerowitz, 1991; Theissen, 2001). Although the MADS box gene family originated prior to the divergence between major eukaryotic lineages (Gramzow et al., 2010), extensive duplication events have resulted in the expansion of MADS box genes involved in floral development (Purugganan et al., 1995; Kramer et al., 1998, 2006; Causier et al., 2010). Many MADS box paralogs appear to have redundant functions (Vandenbussche et al., 2004; de Martino et al., 2006; Geuten and Irish, 2010), but others have clearly diverged in function (Airoldi and Davies, 2012). Some examples include the divergence of the *AP3/TM6* paralogs in petunia (*Petunia hybrida*) due to a change in regulatory elements (Rijpkema et al., 2006), the modification of preexisting floral regulators leading to the development of novel floral organ types in the Ranunculales (Kramer et al., 2007; Rasmussen et al., 2009), and the lineage-specific expansion and divergence of the *DEFICIENS* genes in orchids (Orchidaceae spp.), giving rise to unique floral features (Mondragón-Palomino and Theissen, 2011). These examples highlight the importance of duplication and divergence of floral identity genes as a major contributor to the evolution of floral morphology. This process also has likely been important to the evolution of vegetative organs. For example, a duplicated KNOX transcription factor has acquired a novel regulatory pattern that regulates leaf shape and aboveground architecture in plants (Furumizu et al., 2015). Finally, duplication can contribute to the interactions of plants with other organisms: the duplication of a receptor-like

kinase gene originally involved in mycorrhizal symbiosis gave rise to the lysin motif receptor-like kinase SILYK10 in tomato (*Solanum lycopersicum*), which likely adopted a new role in nodulation with clear adaptive significance (Buendia et al., 2016). These are but a few examples of evolutionary novelties that can be found in the large body of plant genomic and functional studies.

It should be emphasized that a novel molecular function may not necessarily contribute to a novel plant structure/function. In addition, a novel plant structure/function may not necessarily be adaptive. The challenge lies in not only determining the presence of a novel activity but also assessing its adaptive significance. Using the disease resistance (*R*) genes as an example, tandem duplication (Rizzon et al., 2006; Yu et al., 2015) and WGD (Cannon et al., 2002; Plocik et al., 2004; Zhang et al., 2014) have contributed to a net gain of *R* genes over the course of plant evolution. Based on detailed functional studies of selected *R* genes, they are important for eliciting proper host defense responses; thus, their presence is clearly adaptive (Dangl and Jones, 2001; Jones and Dangl, 2006). In addition, strong positive selection driving sequence divergence between some *R* gene duplicates has been noted (Bakker et al., 2006; Ratnaparkhe et al., 2011). Thus, the proliferation of *R* genes has been hypothesized to be an indication of their adaptive value (Holub, 2001). Nonetheless, *R* genes belong to one of the most highly variable gene families (Clark et al., 2007) and tend to have an excess of pseudogenes (Zou et al., 2009a). A substantial number of *R* genes may not be under selection, and for those that are, assessing the adaptive significance is challenging because the biotic factors that interact with the *R* gene products are largely unknown. This challenge is not specific to *R* genes but applies to any duplicate gene thought to contribute to increased fitness.

Ecological Impacts of Duplicate Genes

The evolutionary innovations ascribed to duplicate genes can have important ecological implications. Gene duplication has contributed to developmental novelties that facilitate interactions between plants and other species. For example, the evolution of floral characteristics is strongly associated with functional groups of pollinators (Fenster et al., 2004). The concept of a pollinator syndrome suggests that selective pressure exerted by pollinators results in the convergent evolution of a common set of floral traits. In orchids, the majority of species have only a single pollinator (Tremblay, 1992), but the evolution of specialized morphology for certain pollinators has occurred multiple times (Johnson et al., 1998). According to the orchid code hypothesis, the proximal cause of the diversity of orchid floral structures can be attributed to two DEFICIENS-like transcription factor duplication events followed by gain and/or loss of function in different orchid lineages (Mondragón-Palomino and Theissen, 2008, 2009). Similarly, the duplication and subsequent diversification of CYCLOIDEA2-like transcription factors in Malpighiaceae is thought to have been important for the evolution of bilateral symmetry (zygomorphic) in flowers from radial symmetry (actinomorphic; Zhang et al., 2010).

In addition to developmental novelties that facilitate ecological interactions, duplicate genes are important components in the evolutionary arms race between plants and pathogens/herbivores. Plant defense against herbivores involves specialized metabolites such as the glucosinolates in Brassicales (Demain and Fang, 2000; Halkier and Gershenzon, 2006) and novel glucosinolate pathway components, which likely derived from duplication events (Edger et al., 2015). In particular, the retention of the core Trp pathway gene duplicates derived from the β WGD event appears to have led to the evolution of the glucosinolate pathway (Edger et al., 2015). Similarly, the previously noted function and expansion of *R* gene families are suggestive of the importance of *R* gene duplication in the interactions between plants and pathogens.

Interestingly, aside from plant-nonplant interactions, both *R* genes and specialized metabolic genes also can impact plant speciation. Speciation results from barriers in gene flow, which can arise at different times during development and can involve many different mechanisms, from molecular to environmental (Rieseberg and Willis, 2007). At the genetic level, incompatibility between related species can be explained by the Bateson-Dobzhansky-Muller model, which involves negative epistasis between two or more divergent loci in hybrids (Orr, 1996). After gene duplication, differential loss of function or patterns of subfunctionalization between duplicate loci can, in some situations, result in net loss of function for certain gamete combinations (Force and Lynch, 2000).

Multiple examples of negative interactions resulting from copy number variation and reciprocal silencing of duplicate genes were observed in a survey of speciation genes (Rieseberg and Blackman, 2010). For example, *R* gene duplicates have been hypothesized to play an important role in speciation due to their ability to induce necrosis in hybrids (Bomblies and Weigel, 2007). Defensive compounds such as glucosinolates may function as part of a pollinator syndrome (Demain and Fang, 2000; Halkier and Gershenzon, 2006), which can serve as a reproductive barrier. Hybrid incompatibility has been shown to result from the differential expression of tandemly duplicated receptor-like kinases that are involved in innate immunity in Arabidopsis (Smith et al., 2011) and by negative interactions between tandemly duplicated clusters of receptor-like kinases and subtilisin-like proteases in wild rice (*Oryza rufipogon*) and domesticated rice (Chen et al., 2014).

Because a change in ploidy levels is expected to result in instant reproductive isolation, WGD is seen as a major mechanism of plant speciation (Ramsey and

Schemske, 1998). Consistent with this expectation, there is a significant correlation between the presence of a recent WGD event and the number of extant species in Brassicaceae, Cleomaceae, Fabaceae, Poaceae, and Solanaceae (Soltis et al., 2009). Similarly, an estimated 15% of speciation events in angiosperms and 31% in ferns were associated with an increase in ploidy (Wood et al., 2009). Duplication also has been associated with the evolution of interspecies interaction in plants (Edger et al., 2015) as well as the evolution of novel structures/functions in angiosperms (Soltis and Soltis, 2016), which are thought to be important to speciation and diversification. The radiation of plant species appears to lag significantly behind WGDs, which has led to the proposition that WGD and the subsequent development of novel traits primes a population for speciation by a subsequent dispersal event (Schranz et al., 2012; Tank et al., 2015). Consistent with this model, the timing of recent ploidy events in angiosperms is clustered around the Cretaceous-Paleogene extinction (Vanneste et al., 2014), and WGD events and speciation events in conifers occurred around the time of the Permian-Triassic extinction (Lu et al., 2014; Li et al., 2015b). However, while the lag-modulated association between WGD and diversification has been shown to be significant (Tank et al., 2015), the connections between WGD and species radiation are correlational, not cause and effect. Neopolyploid lineages are more prone to extinction than diploids (Mayrose et al., 2011), which is consistent with the observation that the rate of neopolyploidization in plants is high yet detectable speciation events due to WGD are relatively rare (Otto and Whitton, 2000; Ramsey and Schemske, 2002; Jiao et al., 2011; Moghe and Shiu, 2014). Given that speciation by WGD is rare, the correlation between WGD and radiating events may be because they are unlikely to be observed in association with dispersed events.

Taken together, gene duplications, both small and large scale, may be important in plant adaptation to variable abiotic and biotic environments, in speciation, and in contributing to the diversity of angiosperm species. Although there is indirect evidence, direct demonstration of the role of duplicate genes in ecological adaptation is challenging to come by. In a landmark study examining the genetic basis of how plants adapt to their local environment, 15 fitness quantitative trait loci were identified using an Arabidopsis recombinant inbred population derived from a Swedish accession and an Italian accession (Ågrena et al., 2013). Interestingly, one major quantitative trait locus contains the *C-REPEAT-BINDING FACTOR* (*CBF*) locus, which has three *CBF* genes known to control plant cold tolerance (Gehan et al., 2015). In the Italian population, *CBF2* is nonfunctional, resulting in reduced cold tolerance. Although this is not an example of a new duplicate attaining a novel, adaptive function, it provides direct evidence of the importance of duplicate genes in influencing species range due to abiotic constraints.

Contribution of Duplicate Genes to Agronomic Traits

Gene duplication facilitates the evolution of novel traits that can be subjected not only to natural selection but also to artificial selection, which is important to crop improvement. A recent review has summarized studies revealing that duplicate genes derived from polyploidy can be key to crop domestication and the evolution of stress resistance/tolerance traits (Renny-Byfield and Wendel, 2014). Many crop plants have experienced relatively recent WGDs: 1.5 MYA in *G. hirsutum* (Li et al., 2015a), 13 MYA in soybean (Schmutz et al., 2010), and 0.5 MYA and approximately 3.5 MYA in wheat (Brenchley et al., 2012). In bread wheat, polyploidy contributed to the grain free-threshing character controlled by complex interactions between the duplicate Q transcription factors (Zhang et al., 2011) and to the soft grain character controlled by the *Hardness* locus (Chantret et al., 2005). It is also hypothesized that polyploidization has contributed to the expansion of wheat storage protein genes (Brenchley et al., 2012). In *G. hirsutum*, the recent WGD has resulted in up-regulation and increased selection of fiber genes on the A(t) subgenome compared with the progenitor genes, which is correlated with the production of longer, spinnable fibers (Li et al., 2015a). Similarly, in soybean, nodulation and oil production gene duplicates tend to be retained post WGD (Schmutz et al., 2010) and may contribute to soybean domestication traits. Based on comparative genomic evidence, it is also argued that selection in *Brassica napus* has led to the preservation of duplicate oil biosynthesis genes (Chalhoub et al., 2014). Although not all of the above examples provide direct evidence, they are suggestive of the impact of gene duplication, particularly polyploidy, on agronomically relevant traits.

In addition to WGD, smaller scale duplications can contribute to agronomic traits. Given that a number of genes involved in plant stress tolerance or resistance are found in tandem clusters and display copy number variation (Ellis et al., 2000), it is expected that tandem duplication has played a major role in these traits. For example, the NBS-LRR gene family, whose members have roles in biotic stress resistance, are highly variable between rice cultivars (Yang et al., 2008). In terms of abiotic stress tolerance, one prominent example is the rice *Submergence1* (*Sub1*) tandem cluster, which contains multiple ethylene-responsive factor genes (*Sub1A*, *Sub1B*, and *Sub1C*) and is involved in submergence tolerance (Fukao et al., 2006). Through comparisons of rice cultivars and wild rice species, it was shown that the Sub1A haplotype likely arose recently, potentially after rice domestication (Fukao et al., 2009).

Tandem duplicates also are implicated in the evolution of other agronomic traits. For example, copy number variation in a tandem duplicated region controls rice grain size diversity, an important quality trait (Wang et al., 2015b). This region contains two copies of *Grain Length on Chromosome7* (*GL7*) that are homologs of the Arabidopsis *LONGIFOLIA* gene, which regulates

cell elongation. Tandem duplication leads to elevated levels of *GL7* expression and increased grain length (Wang et al., 2015b). Based on comparisons of wild and domesticated pepper (*Capsicum annuum*) species, tandem duplicate genes involved in capsaicin biosynthesis have likely contributed to the diversification of pungency in peppers (Qin et al., 2014), which may be the basis of pungency variation among domesticated pepper varieties. Additionally, domesticated tomato provides an example where an agronomic trait was influenced by transposon-mediated duplication. Fruit shape difference between domesticated and wild (*Solanum pimpinellifolium*) tomato is due to the increased expression of *IQD12* after retrotransposon-mediated duplication to a new genomic region (Xiao et al., 2008).

In the previous sections, we have discussed several domestication-related traits that can be attributed to gene duplications. One obvious omission is flowering time. When plants are grown in new environments, the change in photoperiod makes selection for proper flowering time essential. In sunflower (*Helianthus annuus*), the expression of *Flowering Locus T* (*FT*) duplicates is central to the control of flowering time (Blackman et al., 2010). During sunflower domestication, a dominant negative mutation likely occurred in the *FT1* duplicate and contributed to delayed flowering. Another example is the *Heading Date1* (*HD1*) locus, which encodes a member of the CONSTANS transcription factor family (Liu et al., 2015). In sorghum (*Sorghum bicolor*), foxtail millet (*Setaria italica*), and rice, mutant *HD1* orthologs have contributed to delayed flowering time; however, the presence of distinct mutations in each lineage suggests independent origins over the course of parallel domestication (Liu et al., 2015). In summary, because of their abundance and the potential for functional divergence and the acquisition of new functions, duplicate genes have contributed to the evolution of morphological, nutritional, and physiological traits in crops.

FUTURE DIRECTIONS

Studies based on accumulating comparative and functional genomic data have contributed to our understanding of the life cycle of duplicated genes, including their origins, longevity, mechanisms of retention, molecular functional implications, and impacts on plant evolution and ecology. Nonetheless, there are still many unresolved questions (see "Outstanding Questions"). Although metrics like half-life provide a good indicator of the average behavior of duplicate genes, there is large variance in duplicate longevity (Lynch and Conery, 2003; Maere et al., 2005). Why are some duplicates retained longer than predicted by the average half-life? What are the factors contributing to short-lived duplicates? The answers to these questions lie in a better understanding of retention mechanisms, including neofunctionalization, dosage effect, EAC, DDC, gene balance, and paralog interference. In particular, the challenge is to determine the relative contributions of these retention mechanisms. Another challenge is that knowledge of functional divergence alone is insufficient to distinguish between retention mechanisms. Knowledge of ancestral functions and expression state, which can only be inferred, also is required.

The retention mechanisms outlined above all involve natural selection on existing and/or novel functions. Thus, there must be a fitness cost if one of the duplicates is lost. There are but a few studies that have directly addressed the fitness contribution of duplicates (DeLuna et al., 2008; Qian and Zhang, 2014). Instead, the great majority of studies on plant duplicate genes have focused on morphological, developmental, and/or physiological phenotypes of loss-of-function mutants in highly controlled environments. In many cases, the lack of a phenotype when one duplicate is lost is attributed to genetic redundancy (Hanada et al., 2009a). Although genetic redundancy is an authentic phenomenon, one cannot rule out the possibility that the specific environment requiring the function of the apparently redundant duplicate has not yet been identified. It is also possible that there are subtle phenotypes that remain undetected but have significant fitness consequences (Ågrena et al., 2013). Testing these two possibilities requires assessing the fitness cost of loss-of-function mutations in the field, preferably in native environments. Recent studies of Arabidopsis local adaptation show that it is feasible to detect minute fitness effects (Ågrena et al., 2013). In addition, especially for recently duplicated genes (i.e. from duplication events that occurred approximately 1 MYA), one cannot rule out the possibility that some duplicates persist because not enough time has passed for them to be removed by genetic drift, even though they are no longer functional (Kimura and Ohta, 1969). This is particularly true for selfing plants like Arabidopsis, where even deleterious alleles are not efficiently eliminated (Bustamante et al., 2002). Given the diversity in life histories and environments encountered by different plant species, knowledge of the past history (i.e. history of selection, bottlenecks, gene flow, changes in effective population size, and mating system) will be necessary to better estimate the contribution of drift to the persistence of duplicates. This knowledge can be partially acquired from comparative studies of variation in duplicate gene content within and between related plant species. Initiatives like the Arabidopsis 1001 Genomes Project (Cao et al., 2011) and parallel efforts in other species may soon provide some insights in this regard.

Genome-wide studies of duplicate genes have revealed that retained duplicates tend to have particular patterns at the sequence, expression, and molecular function levels (Ganko et al., 2007; Carretero-Paulet and Fares, 2012; Jiang et al., 2013). However, each pattern can only marginally predict duplicate retention (Jiang et al., 2013; Moghe et al., 2014), and there remains much unexplained variance even when most factors that have been shown to be correlated with retention are combined

OUTSTANDING QUESTIONS

- Why are some duplicates retained for longer periods of time than expected, and what factors contribute to differences in duplicate gene longevity?
- To what extent are duplicate genes functionally redundant? Is pure genetic redundancy possible?
- What are the systems-level impacts of gene duplication? Are these impacts dependent on duplication mechanism?
- What are the relative contributions of retention mechanisms, singly or in combination, to the maintenance of duplicate pairs?
- What is the adaptive significance of novel structures and functions that are derived from gene duplication?
- How has gene duplication contributed to the evolution of new species, particularly those of ecological, evolutionary, and agricultural significance?

in a single predictive framework (Jiang et al., 2013; Moghe et al., 2014). This suggests that additional, unknown factors may contribute to duplicate retention. Considering that the function of a gene is directly or indirectly influenced by many other genes, the retention of a duplicate gene should be influenced by other genes that are closely linked to the duplicate in the gene network (i.e. in the same functional module), a possibility consistent with the expected outcome of the gene-balance model (Birchler et al., 2005; Birchler and Veitia, 2007, 2010). This network idea can be pushed to an even higher level of organization by asking how other modules influence the retention of duplicate genes in an entire module. Thus, a systems-level understanding (i.e. knowledge of the architecture of the gene networks and the nature of the connections between genes) is essential to assess how duplicates are influenced by other network components. This knowledge will be helpful not only for addressing the rather esoteric question of how duplicate genes are retained but also for understanding how duplicate genes collectively influence molecular functions, physiology, and development in plants. For example, due to the high rate of transcription factor retention, a gene is not only regulated by transcription factors from different families but also bound by multiple members of the same family (Macneil and Walhout, 2011). The expression level of the gene in question is thus determined by a large number of duplicate factors. Without a systems-level understanding of how duplicates differ in their functions, it will not be possible to gain a complete picture of how a gene is regulated.

Ultimately, our interest in studying duplicate genes lies in their evolutionary, ecological, and agronomic impacts. Although the acquisition of novel molecular functions among duplicates is common (Blanc and Wolfe, 2004b), this alone is not sufficient to conclude that there was an impact on evolution. An apparently novel function (e.g. a new biochemical activity, pattern of expression, or interaction) with no effect on fitness could have been fixed by genetic drift. When a duplicate pair is retained because of dosage sensitivity and paralog interference, an unrelated change in function may be misidentified as a functional novelty. Furthermore, apparent novelty at the single gene level may result from larger scale changes in the genome following WGD, such as fractionation (Schnable et al., 2011). To be able to claim that natural selection is important, either a direct study of fitness or molecular evidence of a nonneutral mutation is required. Similarly, the contribution of duplication to the diversification of a multitude of plant traits has been explored, but few have examined the adaptive significance of those traits (Ågrena et al., 2013; Gehan et al., 2015). More examples illustrating the impact of novel traits on plant adaptation are needed.

Understanding the impact of duplicated genes is important in light of the challenges facing agriculture in the 21st century, including both the old problems of yield, disease resistance, and stress tolerance as well as new issues related to global climate change. Addressing the grand challenge of food security will not only require improving our ability to modify plant traits (Halpin, 2005) but also our ability to identify the causative loci of desirable traits (Mickelbart et al., 2015) and the genomic context in which they exist (Vaughan et al., 2007). In this regard, a continuing effort to understand how duplicate genes have contributed to novel functions, expansion of gene families, and the structure of the genome as a whole is necessary. Considering how duplicate genes have contributed to evolutionary novelties and diversity in plants, understanding the evolution of duplicate gene functions holds the key to understanding the future of both natural and domesticated populations, particularly in light of impending environmental shift due to global climate change.

Received April 2, 2016; accepted May 17, 2016; published June 10, 2016.

LITERATURE CITED

Abdelsamad A, Pecinka A (2014) Pollen-specific activation of *Arabidopsis* retrogenes is associated with global transcriptional reprogramming. Plant Cell **26:** 3299–3313

Ågrena J, Oakley CG, McKay JK, Lovell JT, Schemske DW (2013) Genetic mapping of adaptation reveals fitness tradeoffs in Arabidopsis thaliana. Proc Natl Acad Sci USA **110:** 21077–21082

Airoldi CA, Davies B (2012) Gene duplication and the evolution of plant MADS-box transcription factors. J Genet Genomics **39:** 157–165

Akhunov ED, Sehgal S, Liang H, Wang S, Akhunova AR, Kaur G, Li W, Forrest KL, See D, Simková H, et al (2013) Comparative analysis of syntenic genes in grass genomes reveals accelerated rates of gene structure and coding sequence evolution in polyploid wheat. Plant Physiol **161:** 252–265

Aklilu BB, Culligan KM (2016) Molecular evolution and functional diversification of Replication Protein A1 in plants. Front Plant Sci **7:** 33

Aklilu BB, Soderquist RS, Culligan KM (2014) Genetic analysis of the Replication Protein A large subunit family in Arabidopsis reveals unique and overlapping roles in DNA repair, meiosis and DNA replication. Nucleic Acids Res **42:** 3104–3118

Alleman M, Freeling M (1986) The Mu transposable elements of maize: evidence for transposition and copy number regulation during development. Genetics **112:** 107–119

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res **25:** 3389–3402

Alvarez-Ponce D, Fares MA (2012) Evolutionary rate and duplicability in the Arabidopsis thaliana protein-protein interaction network. Genome Biol Evol **4:** 1263–1274

Ambrosino L, Bostan H, di Salle P, Sangiovanni M, Vigilante A, Chiusano ML (2016) pATsi: paralogs and singleton genes from Arabidopsis thaliana. Evol Bioinform Online **12:** 1–7

Amoutzias GD, Robertson DL, Van de Peer Y, Oliver SG (2008) Choose your partners: dimerization in eukaryotic transcription factors. Trends Biochem Sci **33:** 220–229

Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. Nature **408:** 796–815

Arabidopsis Interactome Mapping Consortium (2011) Evidence for network evolution in an Arabidopsis interactome map. Science **333:** 601–607

Arsovski AA, Pradinuk J, Guo XQ, Wang S, Adams KL (2015) Evolution of cis-regulatory elements and regulatory networks in duplicated genes of Arabidopsis. Plant Physiol **169:** 2982–2991

Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, et al (2000) Gene Ontology: tool for the unification of biology. Nat Genet **25:** 25–29

Bailey JA, Gu Z, Clark RA, Reinert K, Samonte RV, Schwartz S, Adams MD, Myers EW, Li PW, Eichler EE (2002) Recent segmental duplications in the human genome. Science **297:** 1003–1007

Bailey JA, Liu G, Eichler EE (2003) An Alu transposition model for the origin and expansion of human segmental duplications. Am J Hum Genet **73:** 823–834

Baker CR, Hanson-Smith V, Johnson AD (2013) Following gene duplication, paralog interference constrains transcriptional circuit evolution. Science **342:** 104–108

Bakker EG, Toomajian C, Kreitman M, Bergelson J (2006) A genome-wide survey of R gene polymorphisms in *Arabidopsis*. Plant Cell **18:** 1803–1818

Beilstein MA, Nagalingum NS, Clements MD, Manchester SR, Mathews S (2010) Dated molecular phylogenies indicate a Miocene origin for Arabidopsis thaliana. Proc Natl Acad Sci USA **107:** 18724–18728

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

Bennetzen JL (2005) Transposable elements, gene creation and genome rearrangement in flowering plants. Curr Opin Genet Dev **15:** 621–627

Benovoy D, Drouin G (2006) Processed pseudogenes, processed genes, and spontaneous mutations in the Arabidopsis genome. J Mol Evol **62:** 511–522

Birchler JA, Riddle NC, Auger DL, Veitia RA (2005) Dosage balance in gene regulation: biological implications. Trends Genet **21:** 219–226

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

Birchler JA, Veitia RA (2010) The gene balance hypothesis: implications for gene regulation, quantitative traits and evolution. New Phytol **186:** 54–62

Blackman BK, Strasburg JL, Raduski AR, Michaels SD, Rieseberg LH (2010) The role of recently derived FT paralogs in sunflower domestication. Curr Biol **20:** 629–635

Blanc G, Hokamp K, Wolfe KH (2003) A recent polyploidy superimposed on older large-scale duplications in the Arabidopsis genome. Genome Res **13:** 137–144

Blanc G, Wolfe KH (2004a) Widespread paleopolyploidy in model plant species inferred from age distributions of duplicate genes. Plant Cell **16:** 1667–1678

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

Bomblies K, Weigel D (2007) Hybrid necrosis: autoimmunity as a potential gene-flow barrier in plant species. Nat Rev Genet **8:** 382–393

Bowers JE, Chapman BA, Rong J, Paterson AH (2003) Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. Nature **422:** 433–438

Bowman JL, Smyth DR, Meyerowitz EM (1989) Genes directing flower development in *Arabidopsis*. Plant Cell **1:** 37–52

Brenchley R, Spannagl M, Pfeifer M, Barker GL, D'Amore R, Allen AM, McKenzie N, Kramer M, Kerhornou A, Bolser D, et al (2012) Analysis of the bread wheat genome using whole-genome shotgun sequencing. Nature **491:** 705–710

Bridgham JT, Brown JE, Rodríguez-Marí A, Catchen JM, Thornton JW (2008) Evolution of a new function by degenerative mutation in cephalochordate steroid receptors. PLoS Genet **4:** e1000191

Brockington SF, Yang Y, Gandia-Herrero F, Covshoff S, Hibberd JM, Sage RF, Wong GK, Moore MJ, Smith SA (2015) Lineage-specific gene radiations underlie the evolution of novel betalain pigmentation in Caryophyllales. New Phytol **207:** 1170–1180

Brosius J (1991) Retroposons: seeds of evolution. Science **251:** 753

Buendia L, Wang T, Girardin A, Lefebvre B (2016) The LysM receptor-like kinase SlLYK10 regulates the arbuscular mycorrhizal symbiosis in tomato. New Phytol **210:** 184–195

Bustamante CD, Nielsen R, Sawyer SA, Olsen KM, Purugganan MD, Hartl DL (2002) The cost of inbreeding in Arabidopsis. Nature **416:** 531–534

Byrne DH, Jelenkovic G (1976) Cytological diploidization in the cultivated octoploid strawberry Fragaria × ananassa. Can J Genet Cytol **18:** 653–659

Cannon SB, Zhu H, Baumgarten AM, Spangler R, May G, Cook DR, Young ND (2002) Diversity, distribution, and ancient taxonomic relationships within the TIR and non-TIR NBS-LRR resistance gene subfamilies. J Mol Evol **54:** 548–562

Cao J, Schneeberger K, Ossowski S, Günther T, Bender S, Fitz J, Koenig D, Lanz C, Stegle O, Lippert C, et al (2011) Whole-genome sequencing of multiple Arabidopsis thaliana populations. Nat Genet **43:** 956–963

Carretero-Paulet L, Fares MA (2012) Evolutionary dynamics and functional specialization of plant paralogs formed by whole and small-scale genome duplications. Mol Biol Evol **29:** 3541–3551

Casneuf T, De Bodt S, Raes J, Maere S, 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

Causier B, Castillo R, Xue Y, Schwarz-Sommer Z, Davies B (2010) Tracing the evolution of the floral homeotic B- and C-function genes through genome synteny. Mol Biol Evol **27:** 2651–2664

Chalhoub B, Denoeud F, Liu S, Parkin IA, Tang H, Wang X, Chiquet J, Belcram H, Tong C, Samans B, et al (2014) Early allopolyploid evolution in the post-Neolithic Brassica napus oilseed genome. Science **345:** 950–953

Chantret N, Salse J, Sabot F, Rahman S, Bellec A, Laubin B, Dubois I, Dossat C, Sourdille P, Joudrier P, et al (2005) Molecular basis of evolutionary events that shaped the hardness locus in diploid and polyploid wheat species (*Triticum* and *Aegilops*). Plant Cell **17:** 1033–1045

Chapman BA, Bowers JE, Feltus FA, Paterson AH (2006) Buffering of crucial functions by paleologous duplicated genes may contribute cyclicality to angiosperm genome duplication. Proc Natl Acad Sci USA **103:** 2730–2735

Chaudhary B, Flagel L, Stupar RM, Udall JA, Verma N, Springer NM, Wendel JF (2009) Reciprocal silencing, transcriptional bias and functional divergence of homeologs in polyploid cotton (Gossypium). Genetics **182:** 503–517

Chen C, Chen H, Lin YS, Shen JB, Shan JX, Qi P, Shi M, Zhu MZ, Huang XH, Feng Q, et al (2014) A two-locus interaction causes interspecific hybrid weakness in rice. Nat Commun **5:** 3357

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

Clark RM, Schweikert G, Toomajian C, Ossowski S, Zeller G, Shinn P, Warthmann N, Hu TT, Fu G, Hinds DA, et al (2007) Common sequence polymorphisms shaping genetic diversity in Arabidopsis thaliana. Science **317:** 338–342

Coen ES, Meyerowitz EM (1991) The war of the whorls: genetic interactions controlling flower development. Nature **353:** 31–37

Conant GC, Birchler JA, Pires JC (2014) Dosage, duplication, and diploidization: clarifying the interplay of multiple models for duplicate gene evolution over time. Curr Opin Plant Biol **19:** 91–98

Conant GC, Wolfe KH (2008) Turning a hobby into a job: how duplicated genes find new functions. Nat Rev Genet **9:** 938–950

Dangl JL, Jones JD (2001) Plant pathogens and integrated defence responses to infection. Nature **411:** 826–833

Dayhoff MO (1976) The origin and evolution of protein superfamilies. Fed Proc **35:** 2132–2138

Dehal P, Boore JL (2005) Two rounds of whole genome duplication in the ancestral vertebrate. PLoS Biol **3:** e314

DeLuna A, Vetsigian K, Shoresh N, Hegreness M, Colón-González M, Chao S, Kishony R (2008) Exposing the fitness contribution of duplicated genes. Nat Genet **40:** 676–681

Demain AL, Fang A (2000) The natural functions of secondary metabolites. Adv Biochem Eng Biotechnol **69:** 1–39

de Martino G, Pan I, Emmanuel E, Levy A, Irish VF (2006) Functional analyses of two tomato APETALA3 genes demonstrate diversification in their roles in regulating floral development. Plant Cell **18:** 1833–1845

Derelle E, Ferraz C, Rombauts S, Rouzé P, Worden AZ, Robbens S, Partensky F, Degroeve S, Echeynié S, Cooke R, et al (2006) Genome analysis of the smallest free-living eukaryote Ostreococcus tauri unveils many unique features. Proc Natl Acad Sci USA **103:** 11647–11652

Des Marais DL, Rausher MD (2008) Escape from adaptive conflict after duplication in an anthocyanin pathway gene. Nature **454:** 762–765

D'Hont A, Denoeud F, Aury JM, Baurens FC, 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:** 213–217

Drouin G, Dover GA (1990) Independent gene evolution in the potato actin gene family demonstrated by phylogenetic procedures for resolving gene conversions and the phylogeny of angiosperm actin genes. J Mol Evol **31:** 132–150

Du C, Fefelova N, Caronna J, He L, Dooner HK (2009) The polychromatic Helitron landscape of the maize genome. Proc Natl Acad Sci USA **106:** 19916–19921

Duarte JM, Cui L, Wall PK, Zhang Q, Zhang X, Leebens-Mack J, Ma H, Altman N, dePamphilis CW (2006) Expression pattern shifts following duplication indicative of subfunctionalization and neofunctionalization in regulatory genes of Arabidopsis. Mol Biol Evol **23:** 469–478

Durbin ML, McCaig B, Clegg MT (2000) Molecular evolution of the chalcone synthase multigene family in the morning glory genome. Plant Mol Biol **42:** 79–92

Edger PP, Heidel-Fischer HM, Bekaert M, Rota J, Glöckner G, Platts AE, Heckel DG, Der JP, Wafula EK, Tang M, et al (2015) The butterfly plant arms-race escalated by gene and genome duplications. Proc Natl Acad Sci USA **112:** 8362–8366

Ellis J, Dodds P, Pryor T (2000) Structure, function and evolution of plant disease resistance genes. Curr Opin Plant Biol **3:** 278–284

Fenster CB, Armbruster WS, Wilson P, Dudash MR, Thomson JD (2004) Pollination syndromes and floral specialization. Annu Rev Ecol Evol Syst **35:** 375–403

Ferguson AA, Zhao D, Jiang N (2013) Selective acquisition and retention of genomic sequences by Pack-Mutator-like elements based on guanine-cytosine content and the breadth of expression. Plant Physiol **163:** 1419–1432

Field B, Osbourn AE (2008) Metabolic diversification: independent assembly of operon-like gene clusters in different plants. Science **320:** 543–547

Fitch WM (1970) Distinguishing homologous from analogous proteins. Syst Zool **19:** 99–113

Force A, Lynch M (2000) The origin of interspecific genomic incompatibility via gene duplication. Am Nat **165:** 590–605

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

Fornoni J, Ordano M, Pérez-Ishiwara R, Boege K, Domínguez CA (2016) A comparison of floral integration between selfing and outcrossing species: a meta-analysis. Ann Bot (Lond) **117:** 299–306

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:** 110–118

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

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:** 131–139

Fukao T, Harris T, Bailey-Serres J (2009) Evolutionary analysis of the Sub1 gene cluster that confers submergence tolerance to domesticated rice. Ann Bot (Lond) **103:** 143–150

Fukao T, Xu K, Ronald PC, Bailey-Serres J (2006) A variable cluster of ethylene response factor-like genes regulates metabolic and developmental acclimation responses to submergence in rice. Plant Cell **18:** 2021–2034

Furumizu C, Alvarez JP, Sakakibara K, Bowman JL (2015) Antagonistic roles for KNOX1 and KNOX2 genes in patterning the land plant body plan following an ancient gene duplication. PLoS Genet **11:** e1004980

Gallardo MH, Bickham JW, Honeycutt RL, Ojeda RA, Köhler N (1999) Discovery of tetraploidy in a mammal. Nature **401:** 341

Ganko EW, Meyers BC, Vision TJ (2007) Divergence in expression between duplicated genes in Arabidopsis. Mol Biol Evol **24:** 2298–2309

Gehan MA, Park S, Gilmour SJ, An C, Lee CM, Thomashow MF (2015) Natural variation in the C-repeat binding factor cold response pathway correlates with local adaptation of Arabidopsis ecotypes. Plant J **84:** 682–693

Geuten K, Irish V (2010) Hidden variability of floral homeotic B genes in Solanaceae provides a molecular basis for the evolution of novel functions. Plant Cell **22:** 2562–2578

Geuten K, Viaene T, Irish VF (2011) Robustness and evolvability in the B-system of flower development. Ann Bot (Lond) **107:** 1545–1556

Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, Mitros T, Dirks W, Hellsten U, Putnam N, et al (2012) Phytozome: a comparative platform for green plant genomics. Nucleic Acids Res **40:** D1178–D1186

Gramzow L, Ritz MS, Theissen G (2010) On the origin of MADS-domain transcription factors. Trends Genet **26:** 149–153

Graur D, Shuali Y, Li WH (1989) Deletions in processed pseudogenes accumulate faster in rodents than in humans. J Mol Evol **28:** 279–285

Greilhuber J, Borsch T, Müller K, Worberg A, Porembski S, Barthlott W (2006) Smallest angiosperm genomes found in Lentibulariaceae, with chromosomes of bacterial size. Plant Biol (Stuttg) **8:** 770–777

Guo H, Lee TH, Wang X, Paterson AH (2013) Function relaxation followed by diversifying selection after whole-genome duplication in flowering plants. Plant Physiol **162:** 769–778

Guo X, Zhang Z, Gerstein MB, Zheng D (2009) Small RNAs originated from pseudogenes: cis- or trans-acting? PLOS Comput Biol **5:** e1000449

Guo YL (2013) Gene family evolution in green plants with emphasis on the origination and evolution of Arabidopsis thaliana genes. Plant J **73:** 941–951

Haberer G, Hindemitt T, Meyers BC, Mayer KF (2004) Transcriptional similarities, dissimilarities, and conservation of cis-elements in duplicated genes of Arabidopsis. Plant Physiol **136:** 3009–3022

Halkier BA, Gershenzon J (2006) Biology and biochemistry of glucosinolates. Annu Rev Plant Biol **57:** 303–333

Halpin C (2005) Gene stacking in transgenic plants: the challenge for 21st century plant biotechnology. Plant Biotechnol J **3:** 141–155

Hanada K, Kuromori T, Myouga F, Toyoda T, Li WH, Shinozaki K (2009a) Evolutionary persistence of functional compensation by duplicate genes in Arabidopsis. Genome Biol Evol **1:** 409–414

Hanada K, Kuromori T, Myouga F, Toyoda T, Shinozaki K (2009b) Increased expression and protein divergence in duplicate genes is associated with morphological diversification. PLoS Genet **5:** e1000781

Hanada K, Vallejo V, Nobuta K, Slotkin RK, Lisch D, Meyers BC, Shiu SH, Jiang N (2009c) The functional role of pack-MULEs in rice inferred from purifying selection and expression profile. Plant Cell **21:** 25–38

Hanada K, Zou C, Lehti-Shiu MD, Shinozaki K, Shiu SH (2008) Importance of lineage-specific expansion of plant tandem duplicates in the adaptive response to environmental stimuli. Plant Physiol **148:** 993–1003

He C, Saedler H (2005) Heterotopic expression of MPF2 is the key to the evolution of the Chinese lantern of Physalis, a morphological novelty in Solanaceae. Proc Natl Acad Sci USA **102:** 5779–5784

He X, Zhang J (2005) Rapid subfunctionalization accompanied by prolonged and substantial neofunctionalization in duplicate gene evolution. Genetics **169:** 1157–1164

Hittinger CT, Carroll SB (2007) Gene duplication and the adaptive evolution of a classic genetic switch. Nature **449:** 677–681

Holstege FC, Jennings EG, Wyrick JJ, Lee TI, Hengartner CJ, Green MR, Golub TR, Lander ES, Young RA (1998) Dissecting the regulatory circuitry of a eukaryotic genome. Cell **95:** 717–728

Holub EB (2001) The arms race is ancient history in Arabidopsis, the wildflower. Nat Rev Genet **2:** 516–527

Hu G, Koh J, Yoo MJ, Chen S, Wendel JF (2015a) Gene-expression novelty in allopolyploid cotton: a proteomic perspective. Genetics **200:** 91–104

Hu Y, Liang W, Yin C, Yang X, Ping B, Li A, Jia R, Chen M, Luo Z, Cai Q, et al (2015b) Interactions of OsMADS1 with floral homeotic genes in rice flower development. Mol Plant **8:** 1366–1384

Huang R, Hippauf F, Rohrbeck D, Haustein M, Wenke K, Feike J, Sorrelle N, Piechulla B, Barkman TJ (2012) Enzyme functional evolution through improved catalysis of ancestrally nonpreferred substrates. Proc Natl Acad Sci USA **109:** 2966–2971

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

Husband BC, Baldwin SJ, Suda J (2013) The incidence of polyploidy in natural plant populations: major patterns and evolutionary processes. *In* J Greilhuber, J Dolezel, JF Wendel, eds, Plant Genome Diversity. Volume 2. Physical Structure, Behavior and Evolution of Plant Genomes. Springer-Verlag, Vienna, pp 255–276

Innan H, Kondrashov F (2010) The evolution of gene duplications: classifying and distinguishing between models. Nat Rev Genet **11:** 97–108

Jiang N, Bao Z, Zhang X, Eddy SR, Wessler SR (2004) Pack-MULE transposable elements mediate gene evolution in plants. Nature **431:** 569–573

Jiang N, Ferguson AA, Slotkin RK, Lisch D (2011) Pack-Mutator-like transposable elements (Pack-MULEs) induce directional modification of genes through biased insertion and DNA acquisition. Proc Natl Acad Sci USA **108:** 1537–1542

Jiang WK, Liu YL, Xia EH, Gao LZ (2013) Prevalent role of gene features in determining evolutionary fates of whole-genome duplication duplicated genes in flowering plants. Plant Physiol **161:** 1844–1861

Jiao Y, Wickett NJ, Ayyampalayam S, Chanderbali AS, Landherr L, Ralph PE, Tomsho LP, Hu Y, Liang H, Soltis PS, et al (2011) Ancestral polyploidy in seed plants and angiosperms. Nature **473:** 97–100

Jin J, Zhang H, Kong L, Gao G, Luo J (2014) PlantTFDB 3.0: a portal for the functional and evolutionary study of plant transcription factors. Nucleic Acids Res **42:** D1182–D1187

Johnson S, Linder H, Steiner K (1998) Phylogeny and radiation of pollination systems in Disa (Orchidaceae). Am J Bot **85:** 402

Jones JD, Dangl JL (2006) The plant immune system. Nature **444:** 323–329

Kaessmann H, Vinckenbosch N, Long M (2009) RNA-based gene duplication: mechanistic and evolutionary insights. Nat Rev Genet **10:** 19–31

Kaltenegger E, Ober D (2015) Paralogue interference affects the dynamics after gene duplication. Trends Plant Sci **20:** 814–821

Kapitonov VV, Jurka J (2007) Helitrons on a roll: eukaryotic rolling-circle transposons. Trends Genet **23:** 521–529

Kejnovsky E, Leitch IJ, Leitch AR (2009) Contrasting evolutionary dynamics between angiosperm and mammalian genomes. Trends Ecol Evol **24:** 572–582

Kellis M, Birren BW, Lander ES (2004) Proof and evolutionary analysis of ancient genome duplication in the yeast Saccharomyces cerevisiae. Nature **428:** 617–624

Kimura M (1968) Evolutionary rate at the molecular level. Nature **217:** 624–626

Kimura M (1983) The Netural Theory of Molecular Evolution. Cambridge University Press, Cambridge, UK

Kimura M, King JL (1979) Fixation of a deleterious allele at one of two "duplicate" loci by mutation pressure and random drift. Proc Natl Acad Sci USA **76:** 2858–2861

Kimura M, Ohta T (1969) The average number of generations until fixation of a mutant gene in a finite population. Genetics **61:** 763–771

Kitano H (2004) Biological robustness. Nat Rev Genet **5:** 826–837

Kliebenstein DJ, Osbourn A (2012) Making new molecules: evolution of pathways for novel metabolites in plants. Curr Opin Plant Biol **15:** 415–423

Kondrashov F (2010) Gene dosage and duplication. *In* K Dittmar, DA Liberles, eds, Evolution after Gene Duplication. John Wiley & Sons, Hoboken, NJ, pp 215–218

Kramer EM, Dorit RL, Irish VF (1998) Molecular evolution of genes controlling petal and stamen development: duplication and divergence within the APETALA3 and PISTILLATA MADS-box gene lineages. Genetics **149:** 765–783

Kramer EM, Holappa L, Gould B, Jaramillo MA, Setnikov D, Santiago PM (2007) Elaboration of B gene function to include the identity of novel floral organs in the lower eudicot *Aquilegia*. Plant Cell **19:** 750–766

Kramer EM, Su HJ, Wu CC, Hu JM (2006) A simplified explanation for the frameshift mutation that created a novel C-terminal motif in the APETALA3 gene lineage. BMC Evol Biol **6:** 30

Kubatova B, Travnicek P, Bastlova D, Curn V, Jarolimova V, Suda J (2008) DNA ploidy-level variation in native and invasive populations of Lythrum salicaria at a large geographical scale. J Biogeogr **35:** 167–176

Langille MG, Clark DV (2007) Parent genes of retrotransposition-generated gene duplicates in Drosophila melanogaster have distinct expression profiles. Genomics **90:** 334–343

Law M, Childs KL, Campbell MS, Stein JC, Olson AJ, Holt C, Panchy N, Lei J, Jiao D, Andorf CM, et al (2015) Automated update, revision, and quality control of the maize genome annotations using MAKER-P improves the B73 RefGen_v3 gene models and identifies new genes. Plant Physiol **167:** 25–39

Lee TH, Tang H, Wang X, Paterson AH (2013) PGDD: a database of gene and genome duplication in plants. Nucleic Acids Res **41:** D1152–D1158

Lehti-Shiu MD, Shiu SH (2012) Diversity, classification and function of the plant protein kinase superfamily. Philos Trans R Soc Lond B Biol Sci **367:** 2619–2639

Lehti-Shiu MD, Uygun S, Moghe GD, Panchy N, Fang L, Hufnagel DE, Jasicki HL, Feig M, Shiu SH (2015) Molecular evidence for functional divergence and decay of a transcription factor derived from whole-genome duplication in *Arabidopsis thaliana*. Plant Physiol **168:** 1717–1734

Lemos B, Meiklejohn CD, Hartl DL (2004) Regulatory evolution across the protein interaction network. Nat Genet **36:** 1059–1060

Lespinet O, Wolf YI, Koonin EV, Aravind L (2002) The role of lineage-specific gene family expansion in the evolution of eukaryotes. Genome Res **12:** 1048–1059

Li D, Liu Y, Zhong C, Huang H (2010) Morphological and cytotype variation of wild kiwifruit (Actinidia chinensis complex) along an altitudinal and longitudinal gradient in central-west China. Bot J Linn Soc **164:** 72–83

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

Li WH (1983) Evolution of duplicate genes and pseudogenes. *In* M Nei, RK Koehn, eds, Evolution of Genes and Proteins. Sinauer Associates, Sunderland, MA, pp 14–37

Li Z, Baniaga AE, Sessa EB, Scascitelli M, Graham SW, Rieseberg LH, Barker MS (2015b) Early genome duplications in conifers and other seed plants. Sci Adv **1:** e1501084

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:** 326–344

Li Z, Zhang H, Ge S, Gu X, Gao G, Luo J (2009) Expression pattern divergence of duplicated genes in rice. BMC Bioinformatics (Suppl 6) **10:** S8

Lisch D (2013) How important are transposons for plant evolution? Nat Rev Genet **14:** 49–61

Lister R, O'Malley RC, Tonti-Filippini J, Gregory BD, Berry CC, Millar AH, Ecker JR (2008) Highly integrated single-base resolution maps of the epigenome in Arabidopsis. Cell **133:** 523–536

Liu H, Liu H, Zhou L, Zhang Z, Zhang X, Wang M, Li H, Lin Z (2015) Parallel domestication of the heading date 1 gene in cereals. Mol Biol Evol **32:** 2726–2737

Liu SL, Baute GJ, Adams KL (2011) Organ and cell type-specific complementary expression patterns and regulatory neofunctionalization between duplicated genes in Arabidopsis thaliana. Genome Biol Evol **3:** 1419–1436

Lloyd JP, Seddon AE, Moghe GD, Simenc MC, Shiu SH (2015) Characteristics of plant essential genes allow for within- and between-species prediction of lethal mutant phenotypes. Plant Cell **27:** 2133–2147

Lu F, Lipka AE, Glaubitz J, Elshire R, Cherney JH, Casler MD, Buckler ES, Costich DE (2013) Switchgrass genomic diversity, ploidy, and evolution: novel insights from a network-based SNP discovery protocol. PLoS Genet **9:** e1003215

Lu Y, Ran JH, Guo DM, Yang ZY, Wang XQ (2014) Phylogeny and divergence times of gymnosperms inferred from single-copy nuclear genes. PLoS ONE **9:** e107679

Lynch M, Conery JS (2000) The evolutionary fate and consequences of duplicate genes. Science **290:** 1151–1155

Lynch M, Conery JS (2003) The evolutionary demography of duplicate genes. J Struct Funct Genomics **3:** 35–44

Lyons E, Freeling M (2008) How to usefully compare homologous plant genes and chromosomes as DNA sequences. Plant J **53:** 661–673

Lyons E, Pedersen B, Kane J, Alam M, Ming R, Tang H, Wang X, Bowers J, Paterson A, Lisch D, et al (2008) Finding and comparing syntenic

regions among Arabidopsis and the outgroups papaya, poplar, and grape: CoGe with rosids. Plant Physiol **148:** 1772–1781

Lysak MA, Koch MA, Pecinka A, Schubert I (2005) Chromosome triplication found across the tribe Brassiceae. Genome Res **15:** 516–525

Ma Y, Wang J, Zhong Y, Geng F, Cramer GR, Cheng ZM (2015) Subfunctionalization of cation/proton antiporter 1 genes in grapevine in response to salt stress in different organs. Hortic Res **2:** 15031

Macneil LT, Walhout AJ (2011) Gene regulatory networks and the role of robustness and stochasticity in the control of gene expression. Genome Res **21:** 645–657

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

Maere S, Van de Peer Y (2010) Duplicate retention after small- and large-scale duplications. *In* K Dittmar, DA Liberles, eds, Evolution after Gene Duplication. John Wiley & Sons, Hoboken, NJ, pp 31–56

Makino T, McLysaght A (2012) Positionally biased gene loss after whole genome duplication: evidence from human, yeast, and plant. Genome Res **22:** 2427–2435

Mayrose I, Zhan SH, Rothfels CJ, Magnuson-Ford K, Barker MS, Rieseberg LH, Otto SP (2011) Recently formed polyploid plants diversify at lower rates. Science **333:** 1257

McAdams HH, Arkin A (1999) It's a noisy business! Genetic regulation at the nanomolar scale. Trends Genet **15:** 65–69

McCarthy EW, Mohamed A, Litt A (2015) Functional divergence of APETALA1 and FRUITFULL is due to changes in both regulation and coding sequence. Front Plant Sci **6:** 1076

Michaels SD, Bezerra IC, Amasino RM (2004) FRIGIDA-related genes are required for the winter-annual habit in Arabidopsis. Proc Natl Acad Sci USA **101:** 3281–3285

Mickelbart MV, Hasegawa PM, Bailey-Serres J (2015) Genetic mechanisms of abiotic stress tolerance that translate to crop yield stability. Nat Rev Genet **16:** 237–251

Mizutani M, Ohta D (2010) Diversification of P450 genes during land plant evolution. Annu Rev Plant Biol **61:** 291–315

Moghe GD, Hufnagel DE, Tang H, Xiao Y, Dworkin I, Town CD, Conner JK, Shiu SH (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

Moghe GD, Last RL (2015) Something old, something new: conserved enzymes and the evolution of novelty in plant specialized metabolism. Plant Physiol **169:** 1512–1523

Moghe GD, Shiu SH (2014) The causes and molecular consequences of polyploidy in flowering plants. Ann N Y Acad Sci **1320:** 16–34

Mondragón-Palomino M, Theissen G (2008) MADS about the evolution of orchid flowers. Trends Plant Sci **13:** 51–59

Mondragón-Palomino M, Theissen G (2009) Why are orchid flowers so diverse? Reduction of evolutionary constraints by paralogues of class B floral homeotic genes. Ann Bot (Lond) **104:** 583–594

Mondragón-Palomino M, Theissen G (2011) Conserved differential expression of paralogous DEFICIENS- and GLOBOSA-like MADS-box genes in the flowers of Orchidaceae: refining the 'orchid code.' Plant J **66:** 1008–1019

Moore MJ, Bell CD, Soltis PS, Soltis DE (2007) Using plastid genome-scale data to resolve enigmatic relationships among basal angiosperms. Proc Natl Acad Sci USA **104:** 19363–19368

Murat F, Van de Peer Y, Salse J (2012) Decoding plant and animal genome plasticity from differential paleo-evolutionary patterns and processes. Genome Biol Evol **4:** 917–928

Myburg AA, Grattapaglia D, Tuskan GA, Hellsten U, Hayes RD, Grimwood J, Jenkins J, Lindquist E, Tice H, Bauer D, et al (2014) The genome of Eucalyptus grandis. Nature **510:** 356–362

Nourmohammad A, Lässig M (2011) Formation of regulatory modules by local sequence duplication. PLOS Comput Biol **7:** e1002167

Nowak MA, Boerlijst MC, Cooke J, Smith JM (1997) Evolution of genetic redundancy. Nature **388:** 167–171

Nützmann HW, Osbourn A (2014) Gene clustering in plant specialized metabolism. Curr Opin Biotechnol **26:** 91–99

Ohno S (1970) Evolution by Gene Duplication. Springer-Verlag, New York

Orr HA (1996) Dobzhansky, Bateson, and the genetics of speciation. Genetics **144:** 1331–1335

Otto SP, Whitton J (2000) Polyploid incidence and evolution. Annu Rev Genet **34:** 401–437

Panopoulou G, Hennig S, Groth D, Krause A, Poustka AJ, Herwig R, Vingron M, Lehrach H (2003) New evidence for genome-wide duplications at the origin of vertebrates using an amphioxus gene set and completed animal genomes. Genome Res **13:** 1056–1066

Pellicer J, Fay M, Leitch I (2010) The largest eukaryotic genome of them all. Bot J Linn Soc **164:** 10–15

Plocik A, Layden J, Kesseli R (2004) Comparative analysis of NBS domain sequences of NBS-LRR disease resistance genes from sunflower, lettuce, and chicory. Mol Phylogenet Evol **31:** 153–163

Purugganan MD, Rounsley SD, Schmidt RJ, Yanofsky MF (1995) Molecular evolution of flower development: diversification of the plant MADS-box regulatory gene family. Genetics **140:** 345–356

Qi X, Bakht S, Qin B, Leggett M, Hemmings A, Mellon F, Eagles J, Werck-Reichhart D, Schaller H, Lesot A, et al (2006) A different function for a member of an ancient and highly conserved cytochrome P450 family: from essential sterols to plant defense. Proc Natl Acad Sci USA **103:** 18848–18853

Qian W, Zhang J (2014) Genomic evidence for adaptation by gene duplication. Genome Res **24:** 1356–1362

Qin C, Yu C, Shen Y, Fang X, Chen L, Min J, Cheng J, Zhao S, Xu M, Luo Y, et al (2014) Whole-genome sequencing of cultivated and wild peppers provides insights into Capsicum domestication and specialization. Proc Natl Acad Sci USA **111:** 5135–5140

Ramsey J, Schemske DW (1998) Pathways, mechanisms, and rates of polyploid formation in flowering plants. Annu Rev Ecol Syst **23:** 467–501

Ramsey J, Schemske DW (2002) Neopolyploidy in flowering plants. Annu Rev Ecol Syst **33:** 589–639

Rasmussen DA, Kramer EM, Zimmer EA (2009) One size fits all? Molecular evidence for a commonly inherited petal identity program in Ranunculales. Am J Bot **96:** 96–109

Ratnaparkhe MB, Wang X, Li J, Compton RO, Rainville LK, Lemke C, Kim C, Tang H, Paterson AH (2011) Comparative analysis of peanut NBS-LRR gene clusters suggests evolutionary innovation among duplicated domains and erosion of gene microsynteny. New Phytol **192:** 164–178

Renny-Byfield S, Gallagher JP, Grover CE, Szadkowski E, Page JT, Udall JA, Wang X, Paterson AH, Wendel JF (2014) Ancient gene duplicates in Gossypium (cotton) exhibit near-complete expression divergence. Genome Biol Evol **6:** 559–571

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:** 1063–1071

Renny-Byfield S, Wendel JF (2014) Doubling down on genomes: polyploidy and crop plants. Am J Bot **101:** 1711–1725

Rensing SA, Lang D, Zimmer AD, Terry A, Salamov A, Shapiro H, Nishiyama T, Perroud PF, Lindquist EA, Kamisugi Y, et al (2008) The Physcomitrella genome reveals evolutionary insights into the conquest of land by plants. Science **319:** 64–69

Rhee SY, Wood V, Dolinski K, Draghici S (2008) Use and misuse of the Gene Ontology annotations. Nat Rev Genet **9:** 509–515

Rieseberg LH, Blackman BK (2010) Speciation genes in plants. Ann Bot (Lond) **106:** 439–455

Rieseberg LH, Willis JH (2007) Plant speciation. Science **317:** 910–914

Rijpkema AS, Royaert S, Zethof J, van der Weerden G, Gerats T, Vandenbussche M (2006) Analysis of the petunia TM6 MADS box gene reveals functional divergence within the DEF/AP3 lineage. Plant Cell **18:** 1819–1832

Rizzon C, Ponger L, Gaut BS (2006) Striking similarities in the genomic distribution of tandemly arrayed genes in Arabidopsis and rice. PLOS Comput Biol **2:** e115

Rodgers-Melnick E, Mane SP, Dharmawardhana P, Slavov GT, Crasta OR, Strauss SH, Brunner AM, Difazio SP (2012) Contrasting patterns of evolution following whole genome versus tandem duplication events in Populus. Genome Res **22:** 95–105

Sabara HA, Kron P, Husband BC (2013) Cytotype coexistence leads to triploid hybrid production in a diploid-tetraploid contact zone of Chamerion angustifolium (Onagraceae). Am J Bot **100:** 962–970

Sakai H, Mizuno H, Kawahara Y, Wakimoto H, Ikawa H, Kawahigashi H, Kanamori H, Matsumoto T, Itoh T, Gaut BS (2011) Retrogenes in rice (Oryza sativa L. ssp. japonica) exhibit correlated expression with their source genes. Genome Biol Evol **3:** 1357–1368

Salse J, Bolot S, Throude M, Jouffe V, Piegu B, Quraishi UM, Calcagno T, Cooke R, Delseny M, Feuillet C (2008) Identification and characterization of shared duplications between rice and wheat provide new insight into grass genome evolution. Plant Cell **20:** 11–24

Scannell DR, Wolfe KH (2008) A burst of protein sequence evolution and a prolonged period of asymmetric evolution follow gene duplication in yeast. Genome Res **18:** 137–147

Schlötterer C (2015) Genes from scratch: the evolutionary fate of de novo genes. Trends Genet **31:** 215–219

Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, Nelson W, Hyten DL, Song Q, Thelen JJ, Cheng J, et al (2010) Genome sequence of the palaeopolyploid soybean. Nature **463:** 178–183

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 USA **108:** 4069–4074

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

Schranz ME, Mohammadin S, Edger PP (2012) Ancient whole genome duplications, novelty and diversification: the WGD radiation lag-time model. Curr Opin Plant Biol **15:** 147–153

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

Seoighe C, Wolfe KH (1999) Yeast genome evolution in the post-genome era. Curr Opin Microbiol **2:** 548–554

Sharopova N (2008) Plant simple sequence repeats: distribution, variation, and effects on gene expression. Genome **51:** 79–90

She X, Cheng Z, Zöllner S, Church DM, Eichler EE (2008) Mouse segmental duplication and copy number variation. Nat Genet **40:** 909–914

Shiu SH, Shih MC, Li WH (2005) Transcription factor families have much higher expansion rates in plants than in animals. Plant Physiol **139:** 18–26

Siena LA, Ortiz JP, Calderini O, Paolocci F, Cáceres ME, Kaushal P, Grisan S, Pessino SC, Pupilli F (2016) An apomixis-linked ORC3-like pseudogene is associated with silencing of its functional homolog in apomictic Paspalum simplex. J Exp Bot **67:** 1965–1978

Sikosek T, Chan HS, Bornberg-Bauer E (2012) Escape from adaptive conflict follows from weak functional trade-offs and mutational robustness. Proc Natl Acad Sci USA **109:** 14888–14893

Smith JD, Bickham JW, Gregory TR (2013) Patterns of genome size diversity in bats (order Chiroptera). Genome **56:** 457–472

Smith LM, Bomblies K, Weigel D (2011) Complex evolutionary events at a tandem cluster of Arabidopsis thaliana genes resulting in a single-locus genetic incompatibility. PLoS Genet **7:** e1002164

Soltis DE, Albert VA, Leebens-Mack J, Bell CD, Paterson AH, Zheng C, Sankoff D, Depamphilis CW, Wall PK, Soltis PS (2009) Polyploidy and angiosperm diversification. Am J Bot **96:** 336–348

Soltis DE, Visger CJ, Soltis PS (2014) The polyploidy revolution then…and now: Stebbins revisited. Am J Bot **101:** 1057–1078

Soltis PS, Marchant DB, Van de Peer Y, Soltis DE (2015) Polyploidy and genome evolution in plants. Curr Opin Genet Dev **35:** 119–125

Soltis PS, Soltis DE (2016) Ancient WGD events as drivers of key innovations in angiosperms. Curr Opin Plant Biol **30:** 159–165

Sommer H, Beltrán JP, Huijser P, Pape H, Lönnig WE, Saedler H, Schwarz-Sommer Z (1990) Deficiens, a homeotic gene involved in the control of flower morphogenesis in Antirrhinum majus: the protein shows homology to transcription factors. EMBO J **9:** 605–613

Song K, Lu P, Tang K, Osborn TC (1995) Rapid genome change in synthetic polyploids of Brassica and its implications for polyploid evolution. Proc Natl Acad Sci USA **92:** 7719–7723

Tack DC, Pitchers WR, Adams KL (2014) Transcriptome analysis indicates considerable divergence in alternative splicing between duplicated genes in Arabidopsis thaliana. Genetics **198:** 1473–1481

Takos AM, Knudsen C, Lai D, Kannangara R, Mikkelsen L, Motawia MS, Olsen CE, Sato S, Tabata S, Jørgensen K, et al (2011) Genomic clustering of cyanogenic glucoside biosynthetic genes aids their identification in Lotus japonicus and suggests the repeated evolution of this chemical defence pathway. Plant J **68:** 273–286

Takuno S, Gaut BS (2012) Body-methylated genes in Arabidopsis thaliana are functionally important and evolve slowly. Mol Biol Evol **29:** 219–227

Tank DC, Eastman JM, Pennell MW, Soltis PS, Soltis DE, Hinchliff CE, Brown JW, Sessa EB, Harmon LJ (2015) Nested radiations and the pulse of angiosperm diversification: increased diversification rates often follow whole genome duplications. New Phytol **207:** 454–467

Tautz D, Domazet-Lošo T (2011) The evolutionary origin of orphan genes. Nat Rev Genet **12:** 692–702

Tezuka D, Ito A, Mitsuhashi W, Toyomasu T, Imai R (2015) The rice ent-KAURENE SYNTHASE LIKE 2 encodes a functional ent-beyerene synthase. Biochem Biophys Res Commun **460:** 766–771

Theissen G (2001) Development of floral organ identity: stories from the MADS house. Curr Opin Plant Biol **4:** 75–85

Thibaud-Nissen F, Ouyang S, Buell CR (2009) Identification and characterization of pseudogenes in the rice gene complement. BMC Genomics **10:** 317

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:** 934–946

Throude M, Bolot S, Bosio M, Pont C, Sarda X, Quraishi UM, 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

Tremblay RL (1992) Trends in the pollination ecology of the Orchidaceae: evolution and systematics. Can J Bot **70:** 642–650

True JR, Carroll SB (2002) Gene co-option in physiological and morphological evolution. Annu Rev Cell Dev Biol **18:** 53–80

Tuskan GA, 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

Vandenbussche M, Zethof J, Royaert S, Weterings K, Gerats T (2004) The duplicated B-class heterodimer model: whorl-specific effects and complex genetic interactions in *Petunia hybrida* flower development. Plant Cell **16:** 741–754

Van de Peer Y, Fawcett JA, Proost S, Sterck L, Vandepoele K (2009a) The flowering world: a tale of duplications. Trends Plant Sci **14:** 680–688

Van de Peer Y, Maere S, Meyer A (2009b) The evolutionary significance of ancient genome duplications. Nat Rev Genet **10:** 725–732

van Hoek MJ, Hogeweg P (2007) The role of mutational dynamics in genome shrinkage. Mol Biol Evol **24:** 2485–2494

van Hoek MJ, Hogeweg P (2009) Metabolic adaptation after whole genome duplication. Mol Biol Evol **26:** 2441–2453

Vanin EF (1985) Processed pseudogenes: characteristics and evolution. Annu Rev Genet **19:** 253–272

Vanneste K, Maere S, Van de Peer Y (2014) Tangled up in two: a burst of genome duplications at the end of the Cretaceous and the consequences for plant evolution. Philos Trans R Soc Lond B Biol Sci **369:** 369

Vaughan DA, Balázs E, Heslop-Harrison JS (2007) From crop domestication to super-domestication. Ann Bot (Lond) **100:** 893–901

Veitia RA, Bottani S, Birchler JA (2013) Gene dosage effects: nonlinearities, genetic interactions, and dosage compensation. Trends Genet **29:** 385–393

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

Vukašinović N, Cvrčková F, Eliáš M, Cole R, Fowler JE, Žárský V, Synek L (2014) Dissecting a hidden gene duplication: the Arabidopsis thaliana SEC10 locus. PLoS ONE **9:** e94077

Wang H, Beyene G, Zhai J, Feng S, Fahlgren N, Taylor NJ, Bart R, Carrington JC, Jacobsen SE, Ausin I (2015a) CG gene body DNA methylation changes and evolution of duplicated genes in cassava. Proc Natl Acad Sci USA **112:** 13729–13734

Wang J, Marowsky NC, Fan C (2014a) Divergence of gene body DNA methylation and evolution of plant duplicate genes. PLoS ONE **9:** e110357

Wang S, Adams KL (2015) Duplicate gene divergence by changes in microRNA binding sites in Arabidopsis and Brassica. Genome Biol Evol **7:** 646–655

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

Wang W, Zheng H, Fan C, Li J, Shi J, Cai Z, Zhang G, Liu D, Zhang J, Vang S, et al (2006) High rate of chimeric gene origination by retroposition in plant genomes. Plant Cell **18:** 1791–1802

Wang Y, Tan X, Paterson AH (2013a) Different patterns of gene structure divergence following gene duplication in Arabidopsis. BMC Genomics **14:** 652

Wang Y, Wang X, Lee TH, Mansoor S, Paterson AH (2013b) Gene body methylation shows distinct patterns associated with different gene origins and duplication modes and has a heterogeneous relationship with gene expression in Oryza sativa (rice). New Phytol **198:** 274–283

Wang Y, Xiong G, Hu J, Jiang L, Yu H, Xu J, Fang Y, Zeng L, Xu E, Xu J, et al (2015b) Copy number variation at the GL7 locus contributes to grain size diversity in rice. Nat Genet **47:** 944–948

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

Warren WC, Hillier LW, Marshall Graves JA, Birney E, Ponting CP, Grützner F, Belov K, Miller W, Clarke L, Chinwalla AT, et al (2008) Genome analysis of the platypus reveals unique signatures of evolution. Nature **453:** 175–183

Wolfe KH, Shields DC (1997) Molecular evidence for an ancient duplication of the entire yeast genome. Nature **387:** 708–713

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 USA **106:** 13875–13879

Wray GA, Hahn MW, Abouheif E, Balhoff JP, Pizer M, Rockman MV, Romano LA (2003) The evolution of transcriptional regulation in eukaryotes. Mol Biol Evol **20:** 1377–1419

Xiao H, Jiang N, Schaffner E, Stockinger EJ, van der Knaap E (2008) A retrotransposon-mediated gene duplication underlies morphological variation of tomato fruit. Science **319:** 1527–1530

Yamada K, Lim J, Dale JM, Chen H, Shinn P, Palm CJ, Southwick AM, Wu HC, Kim C, Nguyen M, et al (2003) Empirical analysis of transcriptional activity in the Arabidopsis genome. Science **302:** 842–846

Yang L, Gaut BS (2011) Factors that contribute to variation in evolutionary rate among Arabidopsis genes. Mol Biol Evol **28:** 2359–2369

Yang L, Takuno S, Waters ER, Gaut BS (2011) Lowly expressed genes in Arabidopsis thaliana bear the signature of possible pseudogenization by promoter degradation. Mol Biol Evol **28:** 1193–1203

Yang S, Gu T, Pan C, Feng Z, Ding J, Hang Y, Chen JQ, Tian D (2008) Genetic variation of NBS-LRR class resistance genes in rice lines. Theor Appl Genet **116:** 165–177

Yang X, Tuskan GA, Cheng MZ (2006) Divergence of the Dof gene families in poplar, Arabidopsis, and rice suggests multiple modes of gene evolution after duplication. Plant Physiol **142:** 820–830

Yu J, Ke T, Tehrim S, Sun F, Liao B, Hua W (2015) PTGBase: an integrated database to study tandem duplicated genes in plants. Database (Oxford) **2015:** bav017

Zhang H (2003) Evolution by gene duplication: an update. Trends Ecol Evol **18:** 292

Zhang J (2012) Genetic redundancies and their evolutionary maintenance. Adv Exp Med Biol **751:** 279–300

Zhang R, Murat F, Pont C, Langin T, Salse J (2014) Paleo-evolutionary plasticity of plant disease resistance genes. BMC Genomics **15:** 187

Zhang S, Zhang JS, Zhao J, He C (2015) Distinct subfunctionalization and neofunctionalization of the B-class MADS-box genes in Physalis floridana. Planta **241:** 387–402

Zhang W, Kramer EM, Davis CC (2010) Floral symmetry genes and the origin and maintenance of zygomorphy in a plant-pollinator mutualism. Proc Natl Acad Sci USA **107:** 6388–6393

Zhang Z, Belcram H, Gornicki P, Charles M, Just J, Huneau C, Magdelenat G, Couloux A, Samain S, Gill BS, et al (2011) Duplication and partitioning in evolution and function of homoeologous Q loci governing domestication characters in polyploid wheat. Proc Natl Acad Sci USA **108:** 18737–18742

Zou C, Lehti-Shiu MD, Thibaud-Nissen F, Prakash T, Buell CR, Shiu SH (2009a) Evolutionary and expression signatures of pseudogenes in Arabidopsis and rice. Plant Physiol **151:** 3–15

Zou C, Lehti-Shiu MD, Thomashow M, Shiu SH (2009b) Evolution of stress-regulated gene expression in duplicate genes of Arabidopsis thaliana. PLoS Genet **5:** e1000581