Evolution of Plant Conserved microRNAs After Whole-Genome Duplications

Liuyu Qin [1,2,3], Peng Xu [1,2,4], Yuannian Jiao [1,2,3,4,*]

[1]State Key Laboratory of Plant Diversity and Specialty Crops, Institute of Botany, Chinese Academy of Sciences, Beijing 100093, China
[2]Key Laboratory of Systematic and Evolutionary Botany, Institute of Botany, Chinese Academy of Sciences, Beijing 100093, China
[3]China National Botanical Garden, Beijing 100093, China
[4]University of Chinese Academy of Sciences, Beijing 100049, China

*Corresponding author: E-mail: jiaoyn@ibcas.ac.cn.

Accepted: February 25, 2025

Abstract

MicroRNAs (miRNAs) are a specialized class of small silencing RNAs that regulate gene expression in numerous biological processes in eukaryotes. While the evolutionary dynamics of protein-coding genes after plant whole-genome duplications (WGDs) has been extensively studied, the patterns of evolution for conserved *MIRNAs* (miRNA genes) post-WGDs are less understood. In this study, we systematically investigated miRNAs and their targets in 6 plant species with varying WGD histories. Our findings reveal that WGDs significantly contribute to the expansion of conserved miRNA families. Notably, through homologous analyses of conserved miRNA families, we discovered that beyond the loci derived from WGDs and other duplication events, some conserved miRNA families have independently gained new loci and/or lost syntenic loci in specific lineage or species through evolution. Additionally, our analyses of sequence divergence in conserved miRNAs showed that the mature sequences of miRNA duplicates gradually diverge following WGDs, with this sequence divergence being correlated with that of their adjacent protein-coding genes after recent WGDs. Furthermore, expression and functional divergence analyses of duplicated targets in different miRNA–target interaction scenarios suggest that conserved miRNAs may play crucial roles in regulating the expression of duplicated genes and related regulatory networks following WGDs. In summary, our analyses reveal universal evolutionary patterns of plant conserved miRNAs following WGDs and provide evidence that some miRNA copies in conserved families originated independently during evolution.

Key words: expansion, sequence divergence, co-evolution, de novo origination.

Significance

By comparative analyses of miRNAs in 6 species of monocots and eudicots, we identified expansion and divergence patterns of conserved microRNAs and their corresponding targets following whole-genome duplications and also discovered de novo origination of new copies of these conserved miRNA in certain lineages during evolution. These findings could help us to better understand the origin and evolution of miRNAs in angiosperms.

Introduction

Gene duplications, including whole-genome duplication (WGD), segmental duplication (SD), and tandem duplication (TD), provide essential raw genetic material for evolution (Panchy et al. 2016). WGD, in particular, causes a sudden increase in both genome size and duplicated gene sets, making it a crucial driver of evolutionary novelty and adaptive evolution (Jiao et al. 2011, 2012, 2014; Soltis

and Soltis 2016; Van de Peer et al. 2017; Jiao 2018; Fox et al. 2020). It is widely acknowledged that duplicated genes generally follow one of the 3 evolutionary paths after WGDs: neo-functionalization, sub-functionalization, or non-functionalization (Moore and Purugganan 2005; Roulin et al. 2013; Rensing 2014). The surviving duplicates and their subsequent divergence have the potential to rewire gene regulatory networks, significantly contributing to genetic diversity and, ultimately, biodiversity (De Smet and Van de Peer 2012; Wu et al. 2020; Wang et al. 2021; Wang et al. 2023).

MicroRNAs (miRNAs) are small non-coding RNAs, typically 20 to 24 nucleotides in length. They can recognize their target genes through nearly perfect sequence complementarity, and negatively regulate gene expression post-transcriptionally either by mRNA cleavage or translational suppression (Carthew and Sontheimer 2009; Voinnet 2009; Pasquinelli 2012; Meyers and Axtell 2019; Zhan and Meyers 2023). Numerous studies have reported that miRNAs can play crucial roles in various developmental processes, including patterning of the embryo, meristem development, establishment of lateral organ polarity and boundaries, and the growth of vegetative and reproductive organ (Rubio-Somoza and Weigel 2011; D'Ario et al. 2017; Chen et al. 2018). In addition, significant progress has also been made in understanding the role of miRNAs in plant–environment interactions, such as responses to abiotic/biotic stress, and symbiotic/parasitic interactions (Song et al. 2019; Yu et al. 2019; Zhan and Meyers 2023). Over the past decade, high throughput sequencing technologies and advanced bioinformatics tools have greatly facilitated the identification and functional analyses of miRNAs in numerous plant species. The development of comprehensive databases and tools for miRNA annotation and miRNA–target predictions has provided extensive datasets (Bo and Wang 2005; Kruger and Rehmsmeier 2006; Wu et al. 2012; Axtell 2013b; Dai et al. 2018; Kozomara et al. 2019; Kuang et al. 2019; Guo et al. 2020; Lunardon et al. 2020; Chen et al. 2021; Huang et al. 2022), paving the way for a deeper understanding of miRNA evolution. An increasing number of studies have shown that miRNA loci in plant genomes are in a constant dynamic evolutionary state (Chávez Montes et al. 2014; You et al. 2017).

miRNAs are typically classified into different families based on the nucleotide sequence similarity of their mature forms (Meyers et al. 2008). While sequences within the same miRNA family are very similar, the primary (pri-miRNA) and precursor (pre-miRNA) sequences outside the mature miRNA can be highly variable. A certain set of miRNA families, known as conserved miRNAs, are broadly conserved across most or all sampled land plant lineages. These conserved miRNAs generally have relatively high expression levels, multiple loci, and low sequence variation across diverse plant families, indicating tightly constrained roles and targets. They predominantly target genes involved in fundamental pathways that emerged with or preceded the evolution of land plants or major lineages, such as those involved in plant development and stress responses (Axtell and Bowman 2008; Cuperus et al. 2011; Axtell 2013a). In contrast, the majority of miRNAs are lineage- or species-specific, having originated relatively recently in evolutionary time. These young miRNAs often exhibit low expression abundance and are typically present as single copies. They generally play roles in species-specific processes, such as adaptation to local environments or unique biochemical pathways (Cuperus et al. 2011; Chávez Montes et al. 2014).

Given the high proportion of young miRNAs in plant species, it is proposed that miRNAs have a high birth and death rate (Fahlgren et al. 2007; Axtell and Bowman 2008). Several mechanisms have been suggested for the origin of new miRNAs, including derivation from inverted duplicates of target genes (Allen et al. 2004; He et al. 2014), miniature inverted-repeat transposable elements (MITEs) (Piriyapongsa and Jordan 2008; Li et al. 2011), or random sequences (Felippes et al. 2008; Cui et al. 2017; Baldrich et al. 2018; Guo et al. 2022). Additionally, miRNAs can also originate via extensive genome duplications of existing miRNA families, especially conserved miRNAs (Li and Mao 2007). In the *Arabidopsis thaliana* genome, 18 out of 22 miRNA gene families have been reported to have arisen from WGD, SD, or TD, or a combination of these processes (Maher et al. 2006). For example, the miR166 family, which contains 7 members targeting Class III HD-ZIP genes involved in meristem development, has been reported to have originated from WGD, SD, and TD, followed by divergence in spatial and temporal expression patterns (Maher et al. 2006).

The general evolution patterns of *MIRNAs* after WGDs have been investigated in *Glycine max* (Zhao et al. 2015), *Brassica rapa* (Sun et al. 2015), and several grass species (Abrouk et al. 2012). These studies commonly reveal that retained *MIRNA* duplicates tend to be conserved *MIRNAs* and that *MIRNAs* and their target genes preferentially co-retained or co-lose after WGDs. Additionally, it has been found that *MIRNAs* show biased retention among subgenomes following WGDs, a pattern similar to that observed in protein-coding genes. However, these WGDs occurred relatively recently (approximately 13 million years ago [mya] in *G. max* and 15 mya in *B. rapa*; Schmutz et al. 2010; Wang et al. 2011). Despite these findings, detailed studies on the evolutionary patterns and consequences of conserved *MIRNAs* following WGDs are still lacking. Further investigations are warranted to gain a better understanding of how conserved *MIRNAs* evolve after WGDs.

In this study, we selected pairs of species where one species had experienced an independent WGD after

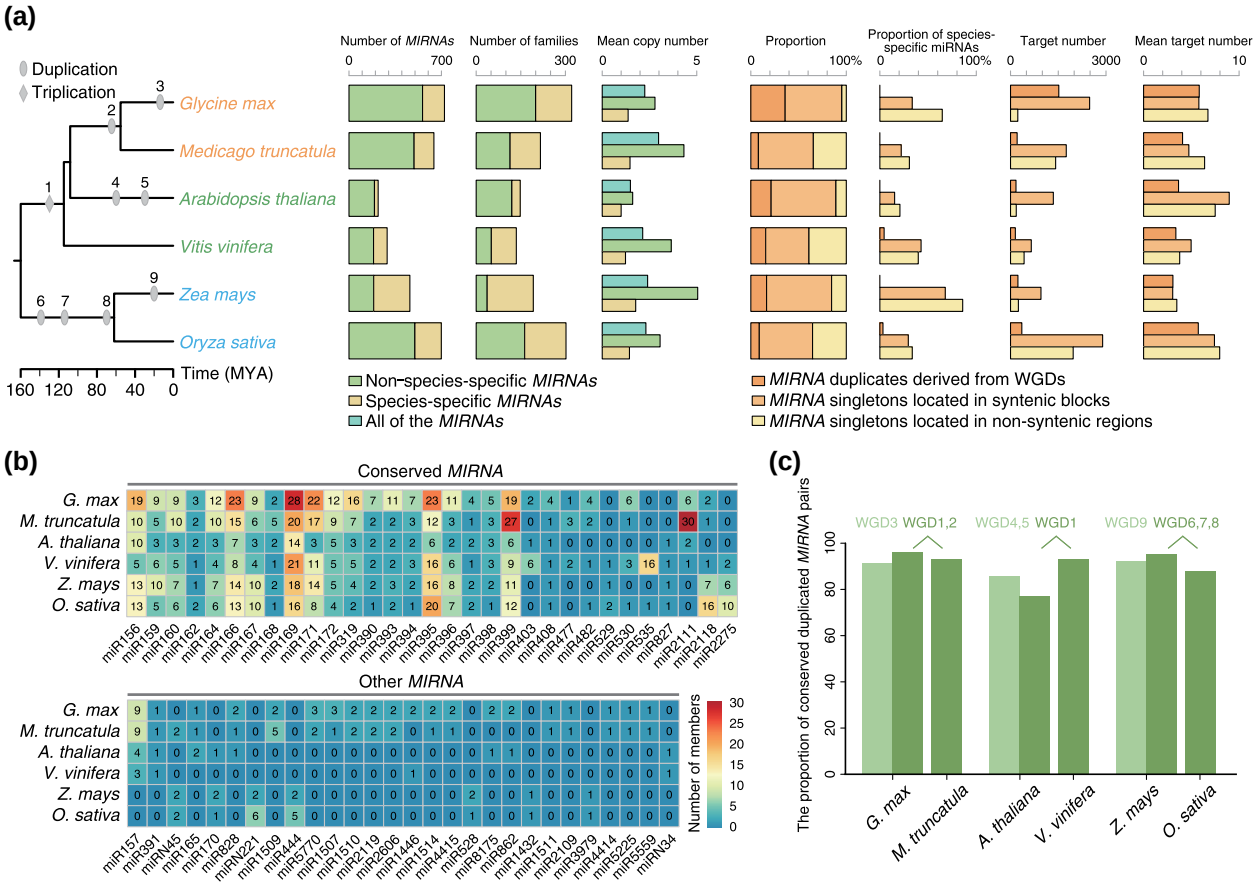

Fig. 1. Overview of the different types of *MIRNAs* and target genes in studied species. a) Detailed classification of *MIRNAs* and identification of target genes: The total number and families of species-specific and non-species-specific *MIRNAs*, as well as the proportion of *MIRNA* duplicates derived from WGDs and *MIRNA* singletons within or outside syntenic blocks are shown for each studied species. The total and average number of target genes for each class of *MIRNAs* are also shown. Each pair of species is represented in the same color. The whole-genome duplication events experienced by each species are indicated on the species tree and assigned numbers. Evolutionary divergence times for each lineage were inferred from TimeTree (http://www.timetree.org/). b) The copy numbers of conserved and 28 selected other miRNA families are displayed for the 6 studied species. c) The percentages of conserved *MIRNAs* in *MIRNA* pairs duplicated from independent and shared WGDs for each studied species.

speciation, and conducted genome-wide, comprehensive studies on the evolution of conserved miRNAs following WGDs. Considering the timing of WGDs, we selected the species pairs *A. thaliana-Vitis vinifera*, *G. max-Medicago truncatula,* and *Zea mays-Oryza sativa*. Our findings indicate that WGDs play a critical role in the expansion of conserved miRNA families. Notably, homologous analyses of conserved miRNA families revealed that members of these families frequently originate and die, and we investigated the mechanisms behind these processes. Additionally, we found that the mature sequences of conserved miRNA can gradually diverge, and that duplicate target genes with different miRNA-target interactions exhibit diverse expression patterns and functional divergence. This study provides new insights into the evolutionary patterns of conserved *MIRNAs* after WGDs. We discovered that these conserved *MIRNAs* can be inherited from ancestral copies or can derive from novel loci in certain lineages or species.

Results

Conserved *MIRNAs* Are More Likely to Retain as Duplicates Than Other *MIRNAs*

To investigate the retention and loss patterns of miRNA following different WGDs in plants, we selected 3 pairs of species with varying WGD histories: *A. thaliana* and *V. vinifera*, *G. max* and *M. truncatula*, and *Z. mays* and *O. sativa* (Lohaus and Van de Peer 2016; Van de Peer et al. 2017; Fig. 1). First, we obtained miRNA annotation information for each species from the plant miRNA database PmiREN (https://pmiren.com/; Guo et al. 2020) and annotated miRNA targets by integrating degradome data and target prediction analyses (supplementary tables S1 to S6, Supplementary Material online). To gain a general understanding of *MIRNA* evolution in these species, we identified the species-specific and non-species-specific *MIRNAs* based on the annotation information (species-specific *MIRNA* are

prefixed by "*MIRN*") and further identified those with or without duplicated copies from WGDs, through intragenomic syntenic analyses. Syntenic *MIRNA* pairs belonging to the same miRNA family were considered as *MIRNA* duplicates derived from WGDs. The remaining *MIRNAs* were considered singletons, including *MIRNAs* located in non-syntenic regions and those in syntenic blocks but without syntenic *MIRNA* pairs. We identified a total of 260, 51, 47, 45, 76 and 62 *MIRNA* duplicates, as well as 463, 592, 174, 244, 385 and 637 *MIRNA* singletons for *G. max*, *M. truncatula*, *A. thaliana*, *V. vinifera*, *Z. mays* and *O. sativa*, respectively (Fig. 1a). The percentage of *MIRNA* duplicates in these species ranged from 7.93% to 35.96%, with species that have undergone independent WGD events (*G. max*, *A. thaliana,* and *Z. mays*) showing a higher proportion of *MIRNA* duplicates compared with their counterparts (supplementary table S7, Supplementary Material online). Furthermore, nearly all *MIRNA* duplicates were non-species-specific (Fig. 1a). These findings suggest that WGDs significantly contribute to the expansion of miRNA families.

To better understand the evolution of conserved *MIRNAs*, we compared the retention and loss of duplicated copies for conserved *MIRNAs* after WGDs with those of other *MIRNAs*. We retrieved 31 conserved miRNA families across land plant species as proposed in previous studies (Chávez Montes et al. 2014; Axtell and Meyers 2018). We then selected the top 28 other miRNA families, which are present in the maximum number of species and have the largest total copy number across all existed species (supplementary table S8, Supplementary Material online), and compared the copy numbers of conserved and these other miRNA families across 6 species. The results show that the copy numbers of most conserved miRNA families are significantly higher than those of other miRNA families. Notably, several conserved miRNA families, such as miR156, miR166, and miR169 exhibit higher copy numbers in species that have undergone recent WGDs, such as *G. max* and *Z. mays* (Fig. 1b). Additionally, we identified *MIRNA* pairs derived from independent WGDs in *G. max* (WGD3), *A. thaliana* (WGD4,5) and *Z. mays* (WGD9), as well as *MIRNA* pairs duplicated from the shared WGDs of each species pair in these 6 species (WGD1,2 for *G. max* and *M. truncatula*, WGD1 for *A. thaliana* and *V. vinifera*, and WGD6,7,8 for *Z. mays* and *O. sativa*) (Fig. 1a and c). For *MIRNA* duplicated pairs derived from recent WGDs in *G. max*, *A. thaliana,* and *Z. mays*, we further selected the pairs with orthologs in their paired species (supplementary tables S9 and S10, Supplementary Material online). We then calculated the percentage of conserved *MIRNAs* in these duplicated pairs and observed that the majority (76.9% to 96.2%) of *MIRNA* duplicates are evolutionary conserved, for *MIRNA* duplicated pairs derived from both recent (independent) and ancient (shared) WGDs in these studied species. These results suggest that conserved *MIRNAs* are more likely to be retained as duplicates after WGDs compared with other *MIRNAs* (Fig. 1c).

Sequence Conservation and Divergence of *MIRNA* Duplicates

Duplicated protein-coding genes often show sequence and/or expression divergence. However, our understanding of sequence divergence in *MIRNA* duplicates remains limited, despite the significant contribution that sequence divergence of mature miRNAs may make to their functional diversity. In this study, we compared the mature sequences of *MIRNA* paralogs and orthologs in 3 pairs of species. For each pair, we first compared the mature sequences of each *MIRNA* duplicate pair in the species that has undergone independent WGDs. Then, we identified the orthologs of the duplicated *MIRNAs* in the other species and compared their mature miRNA sequences (Fig. 2a). For the pair of *A. thaliana* and *V. vinifera*, we first compared paralogous *MIRNA* sequences within *A. thaliana*. Out of 30 duplicated *MIRNA* pairs, 19 pairs (63.3%) had identical mature miRNAs (S1 = S1′), while the remaining 11 pairs (36.7%) exhibit sequence divergence in their mature miRNAs (S1∼S1′). We then compared the sequences of corresponding orthologs in *V. vinifera* with the 2 *MIRNAs* from *A. thaliana*. Among the 19 identical *MIRNA* pairs in *A. thaliana*, we found orthologs in *V. vinifera* for 14 pairs. Of these, 10 had identical mature miRNAs (S1 = S1′=S2), while 4 exhibited sequence divergence (S1 = S1′∼S2). Among the 11 paralogous *MIRNA* pairs in *A. thaliana* that produce divergent mature miRNAs, 8 pairs had orthologs identified in *V. vinifera*. Of these, 7 pairs contained one miRNA that was identical to the corresponding orthologous sequence in *V. vinifera* (S1∼S1′=S2) and the remaining one pair showing divergent ortholog in *V. vinifera* (S1∼S1′∼S2; Fig. 2a, supplementary table S11, Supplementary Material online). Results across the 3 species consistently indicate a high percentage of *MIRNA* duplicate pairs that produce identical mature miRNAs (63.3%, 56.7%, and 68.8% for *A. thaliana*, *G. mays,* and *Z. mays*, respectively). Additionally, most orthologs identified between *A. thaliana* and *V. vinifera,* as well as *Z. mays* and *O. sativa,* demonstrated significant sequence conservation (Fig. 2a).

To explore whether duplicated *MIRNAs* undergo sequence divergence over evolutionary time, we examined *MIRNA* pairs duplicated from various WGDs in *G. max*, *A. thaliana,* and *Z. mays*. We observed that *MIRNA* duplicates with identical mature sequences tend to have a higher percentage (approximately 47% to 70%) of pairs derived from recent WGDs (WGD3, WGD4,5, and WGD9 for *G. max*, *A. thaliana,* and *Z. mays*, respectively) compared with those with diverged mature sequences (around 19% to 40%; Fig. 2b, supplementary table S12, Supplementary

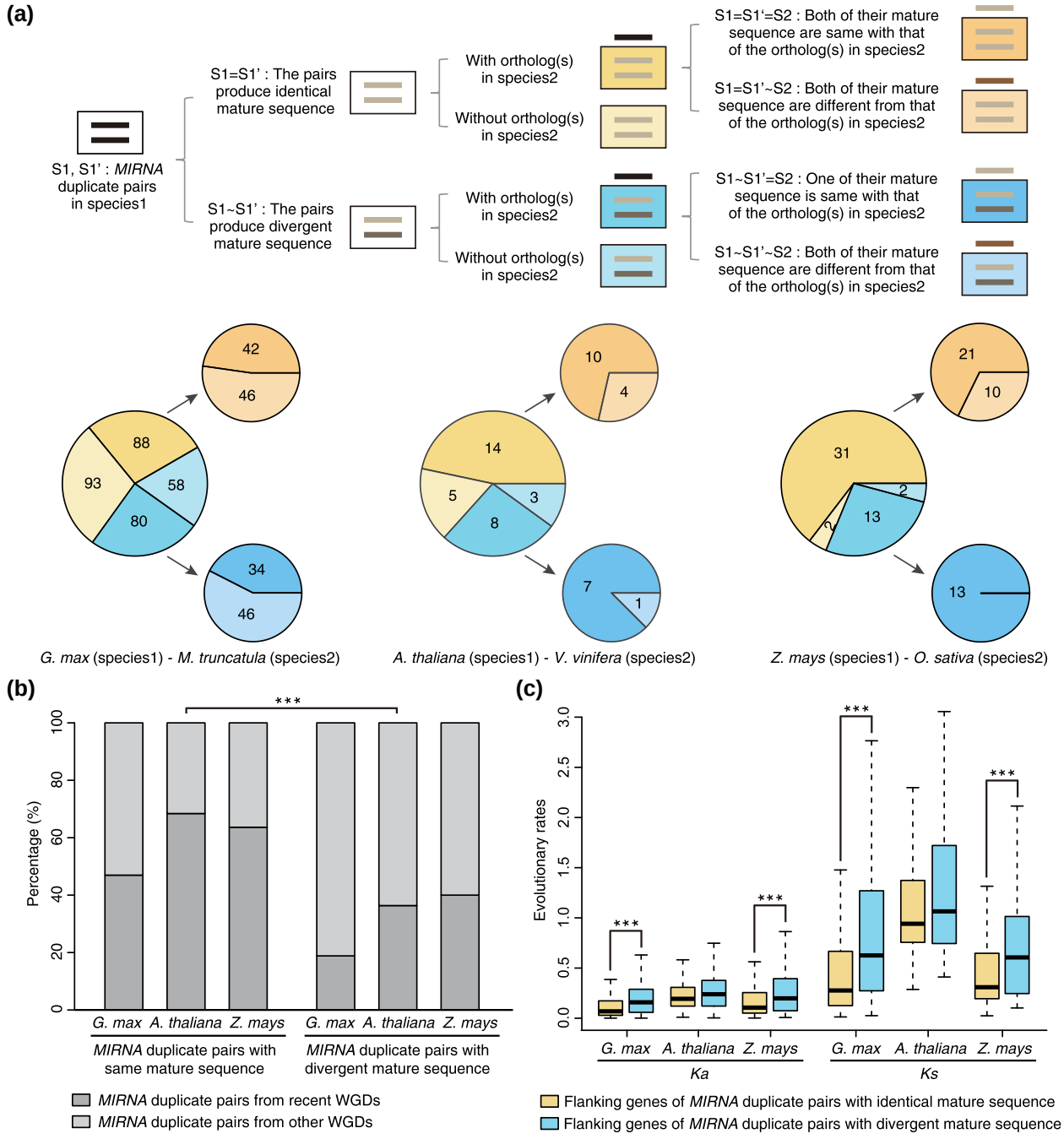

Fig. 2. The sequence divergence of *MIRNA* duplicated pairs and its correlation with the divergence of flanking genes. a) The conservation and divergence of mature sequences in *MIRNA* duplicates, with black figures indicating the number of *MIRNA* duplicates for each scenario. b) The percentage of *MIRNA* duplicates derived from recent and ancient WGDs for paralogous *MIRNA* pairs with identical and divergent mature miRNAs in *Glycine max*, *Arabidopsis thaliana,* and *Zea mays*, respectively. The percentage of *MIRNA* duplicates derived from recent and other WGDs in pairs with identical and divergent mature miRNAs were compared across these 3 species, and the difference between these 2 datasets are significant. The statistical analysis was conducted by Student's *t* test, ***P-value < 0.001. c) Comparison of the evolutionary rates of up- and down-stream 10 flanking genes between *MIRNA* duplicates with identical and divergent mature sequences in *G. max*, *A. thaliana,* and *Z. mays*, respectively. The *Ka* and *Ks* values of syntenic gene pairs flanking *MIRNA* duplicates with identical and divergent mature sequences were grouped and plotted for each species. A significant difference in both *Ka* and *Ks* was observed between miRNA duplicates with identical and divergent mature sequences in *G. max* and *Z. mays* (as shown with asterisk). The statistical analysis was conducted by Student's *t* test, ***P-value < 0.001.

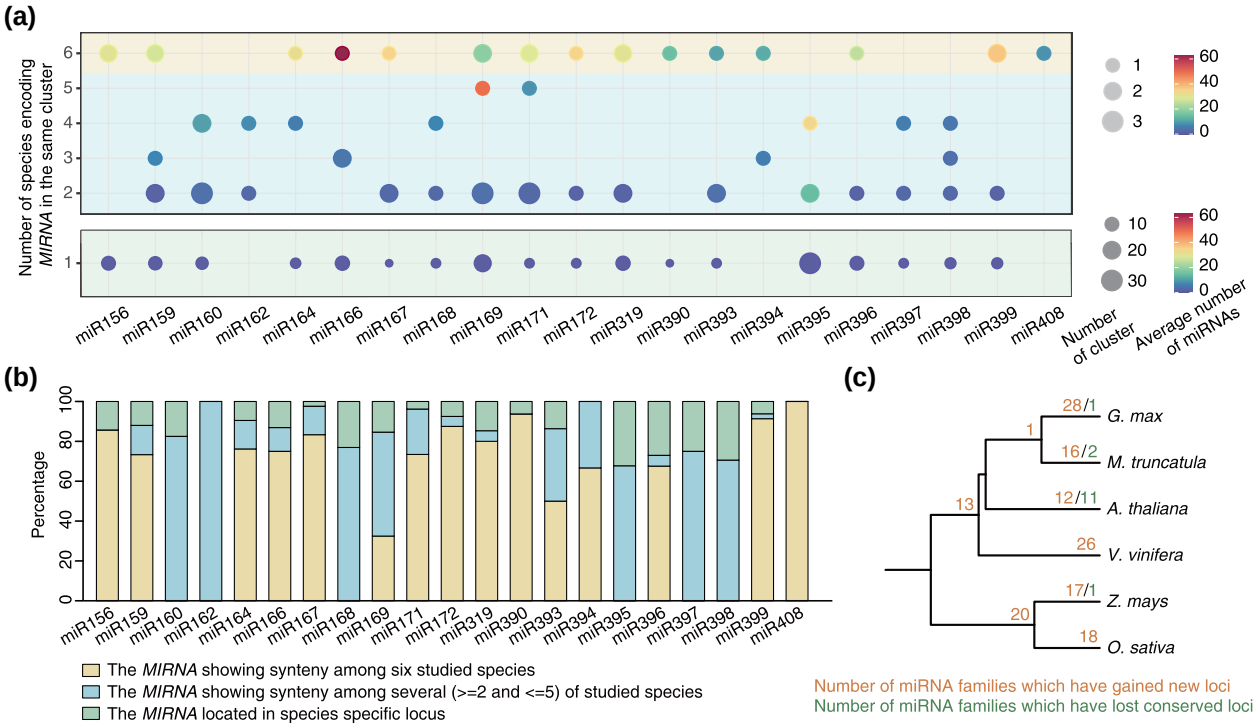

Fig. 3. The birth and death of conserved *MIRNAs* in the 6 studied species are illustrated. a) This section presents the cluster information for 21 conserved miRNA families found in all 6 species examined. The size of each circle represents the number of clusters in different scenarios, while the color indicates the average number of *MIRNA* members within the clusters for each specific scenario. b) This part depicts the percentage of *MIRNA* members with varying degrees of conservation across the 6 species for each conserved miRNA family. c) This section shows the gain and loss of *MIRNA* encoding loci among the 21 conserved miRNA families.

Material online; Student's *t* test, *P*-value <0.001). For instance, in *A. thaliana,* out of 17 *MIRNA* pairs duplicated from the WGD4 or WGD5, 13 pairs had identical mature miRNAs and 4 pairs had diverged mature miRNAs. This accounts for 68.24% (13/19) of all *MIRNA* duplicated pairs with identical mature miRNAs and 36.36% (4/11) of all *MIRNA* duplicated pairs with diverged mature miRNAs (Fig. 2a and b). These findings suggested that the mature sequences of duplicated *MIRNAs* may gradually diverge following WGDs. Additionally, we compared the evolutionary rates of flanking genes of these *MIRNA* pairs. We found that both *Ka* and *Ks* values of the genes flanking duplicated *MIRNAs* with identical mature miRNAs are relatively lower than those of genes flanking duplicated *MIRNAs* with divergent mature miRNAs in all these 3 species, especially in *G. max* and *Z. mays*, which undergone relatively recent WGDs (Fig. 2c; Student's *t* test, *P*-value <0.001). In *M. truncatula*, *V. vinifera*, and *O. sativa*, which have undergone ancient WGDs, this tendency described above is not apparent (supplementary fig. S1, Supplementary Material online). This suggests that the sequence divergence of duplicated *MIRNAs* is consistent with that of their flanking genes after recent WGDs. However, as *MIRNAs* and genes evolve over long periods following WGDs, the correlation between their sequence evolution rates may no longer exist, possibly due to different selection pressure on them.

Expansion of Conserved miRNA Families During Evolution

To better understand the evolution of conserved *MIRNAs*, we identified paralogous and orthologous *MIRNAs* among 6 species based on the interspecific and intraspecific syntenic analyses (supplementary tables S9, S10 and S13, Supplementary Material online). The paralogous and orthologous *MIRNAs* inherited from common conserved ancestral loci were grouped into clusters based on shared members among different syntenic pairs (supplementary table S14, Supplementary Material online). The results show that members of nearly all conserved miRNA families are distributed across multiple clusters, suggesting that most conserved miRNA families have multiple ancestral encoding loci (Fig. 3a). Specifically, for the 21 conserved miRNAs families present in all 6 species, we found that while most have syntenic members among all species, some conserved miRNA families, such as miR160, miR162, and miR168, exhibit syntenic relationships in only a few species (Fig. 3a and b, supplementary table S15, Supplementary Material

online). Moreover, for the conserved miRNA families with syntenic cluster among 6 species, there are also some members that show syntenic relationships only among a few species or specific species (Fig. 3a and b, supplementary figs. S2 to S8, Supplementary Material online). These findings indicate that not all members of the conserved miRNA families display syntenic relationships across the studied species. Possible explanations for this result include translocation or loss of *MIRNAs* in syntenic loci, duplication of *MIRNAs* via mechanisms except WGDs, as well as independent origination of *MIRNAs* in new loci. We conducted a gain and loss analysis of the *MIRNA* loci for each of the 21 conserved miRNA families. Our analysis revealed that independent origination of *MIRNAs* in new loci and the loss of *MIRNAs* derived from the duplication of pre-existing loci occurred frequently within specific lineages or species throughout evolution (Fig. 3c). Moreover, some of these conserved miRNA families have gained multiple new loci in specific lineage or species (supplementary figs. S2 to S8, Supplementary Material online).

To further investigate the possibility of independent origination of *MIRNAs* within conserved families, we analyzed the evolution of *MIRNAs* in specific families. For the conserved miRNA family miR397, we observed that besides the *MIRNAs* showing syntenic relationships among 4 eudicot species and 2 monocot species, each of *G. max*, *A. thaliana,* and *Z. mays* also possesses one *MIRNA* located in non-syntenic encoding loci (Gma-miR397c, Ath-miR397b, and Zma-miR397a; Fig. 4a). Intriguingly, through intergenomic syntenic analysis and sequence alignment, we identified segments with high sequence similarity to Gma-miR397c in the syntenic regions of the *M. truncatula*, *A. thaliana,* and *V. vinifera* genomes (Fig. 4b and c, supplementary figs. S9 and S10, Supplementary Material online). To further confirm the de novo origination of Gma-miR397c, we identified syntenic loci and similar sequences in its closely related species *Phaseolus vulgaris*. The results show that there is also no syntenic encoding locus for Gma-miR397c in *P. vulgaris*, but segment with high sequence similarity to Gma-miR397c was found in corresponding syntenic region. Similarly, we confirmed there is no syntenic encoding locus for Zma-miR397a in its closely related respecies *Sorghum bicolor*. We found segments with high sequence similarity to Zma-miR397a in the syntenic regions of *S. bicolor*, *O. sativa,* and *M. truncatula* genomes (Fig. 4b and c, supplementary fig. S11, Supplementary Material online). The secondary structure prediction of pre-miRNAs revealed that Gma-miR397c and Zma-miR397a can form perfect stem-loop structures. In contrast, most of their similar sequences identified earlier do not exhibit the canonical features of pre-miRNAs (Fig. 4d, supplementary figs. S9 to S11, Supplementary Material online). Based on the principle of maximum parsimony, these observations suggest that the gradual sequence variations may have occasionally led to the formation of hairpin structures or promoter sequences, contributing to the de novo origination of Gma-miR397c, Zma-miR397a, and Ath-miR397b at new encoding loci. Consequently, it can be inferred that some members of conserved miRNA families may have originated independently multiple times throughout evolution, despite producing highly conserved or even identical mature miRNA sequences.

Conserved miRNAs Tend to Target WGD-derived Duplicates

Besides the retention or loss of duplicated gene copies, the diverse expression patterns of paralogous genes significantly contribute to the stabilization and rewiring of biological networks following WGDs. As key regulators of gene expression, miRNAs, especially conserved ones, can play crucial roles in these processes. To assess the regulatory impact of conserved miRNAs on paralogous genes resulting from WGDs, we classified the target genes of miRNAs with varying degrees of conservation into duplicated gene pairs from different duplication mechanisms and singletons across 6 species. We then calculated the percentage of different types of target genes for miRNA at each conservative level, respectively. Similarly, we identified all duplicated genes derived from different duplication mechanisms and singletons and computed the proportion of each gene type among all protein-coding genes of each species. Using the percentage of each type of gene in each genome as references, we found that WGDs-derived target genes make up a higher percentage of total target genes across all 6 species, particularly for conserved miRNAs. This indicates a preference for miRNAs to regulate duplicated genes derived from WGDs, especially conserved ones, in species that have undergone either recent or ancient WGDs (Fig. 5a).

To further investigate the regulatory effects of conserved miRNAs on the paralogous genes and associated biological networks following WGDs, we analyzed the expression and functional divergence of target duplicates with different miRNA regulatory pattern. Firstly, the target genes of conserved *MIRNA* duplicates and singletons in each species were identified and further were classified into duplicates, one member of a duplicated gene pair, and singletons. The target genes and corresponding *MIRNAs* that overlapped through different *MIRNA*–target interaction scenarios were removed (supplementary table S16, Supplementary Material online). We then compared the duplicated status of target genes between conserved *MIRNA* duplicates and singletons in these 6 species. The ratio of target singletons to target duplicates for conserved *MIRNA* duplicates was found to be significantly lower than that for conserved *MIRNA* singletons (Fig. 5b and c; Student's *t* test, *P*-value <0.05). This indicates that targets of *MIRNA* duplicates tend to be retained as duplicates

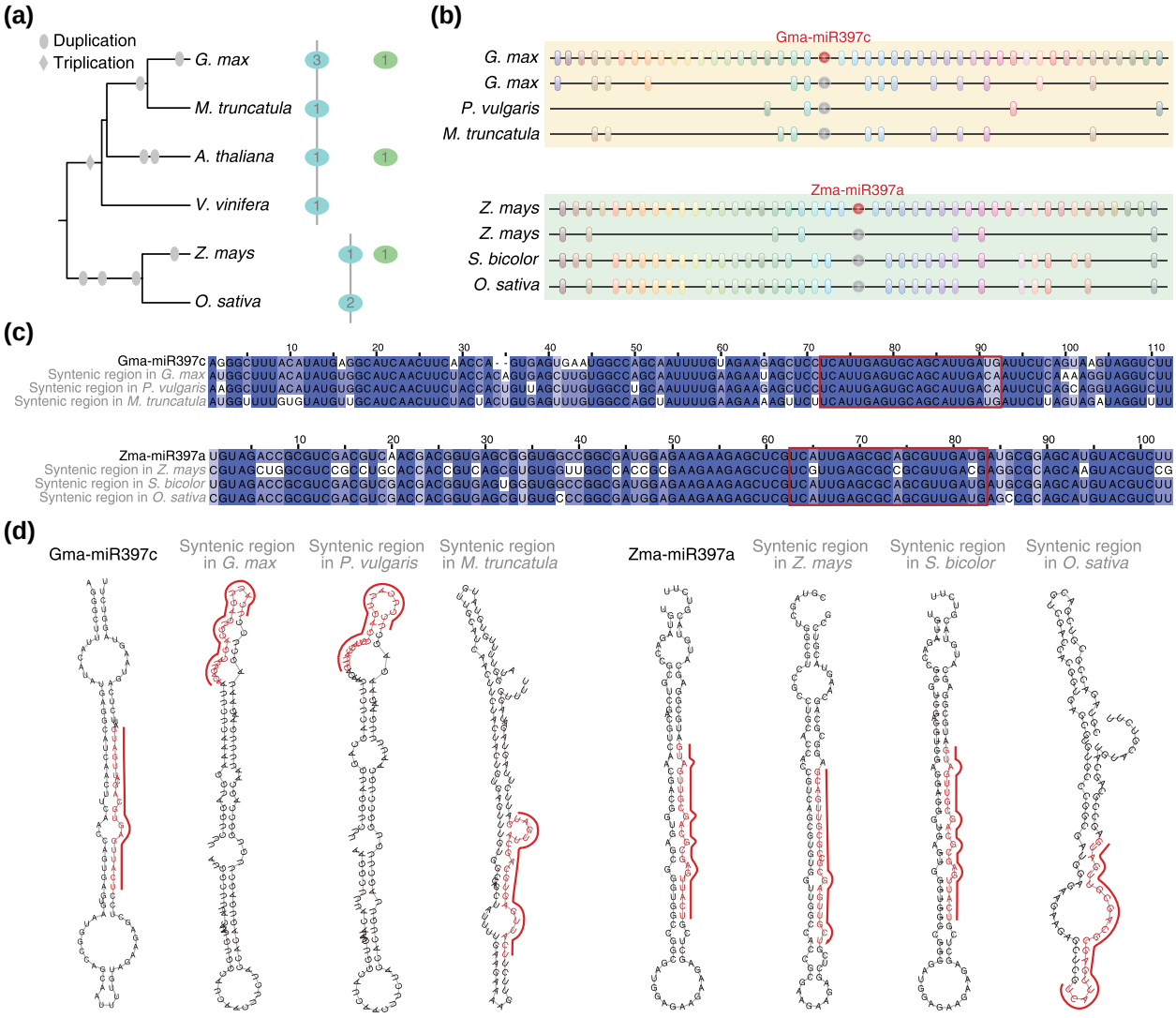

Fig. 4. The independent origination of miR397 members in *Glycine max*, *Arabidopsis thaliana,* and *Zea mays.* a) The syntenic relationship of miR397 among 6 species. Blue ovals linked by lines represent *MIRNAs* inherited from the common ancestral loci across these species, with the numbers inside indicating the count of syntenic *MIRNAs* derived from the corresponding ancestral encoding locus in each species. Green ovals denote *MIRNAs* that originated independently in specific species, with the numbers inside indicating the count of corresponding *MIRNAs* in each species. b) The independent origination of Gma-miR397c and Zma-miR397a is supported by local synteny. Chromosome segments are represented by solid lines. *MIRNAs* are represented by red spots, the identified similar sequences of these 2 *MIRNAs* are represented by grey spots, and paralogous and orthologous genes in the same column are represented by bricks in same colors. Genes with no syntenic matches to the selected regions are not plotted. c) Sequence alignment of Gma-miR397c with its syntenic region in *G. max*, *Phaseolus vulgaris,* and *Medicago truncatula*, and of Zma-miR397a with its syntenic region in *Z. mays*, *Sorghum bicolor,* and *Oryza sativa*. Mature miRNA and their corresponding sequences in syntenic regions are highlighted in red boxes. d) The predicted secondary structure of Gma-miR397c and Zma-miR397a, as well as those of their similar sequences in the corresponding syntenic regions of their own genomes and other studied genomes, respectively. Mature miRNA and their corresponding sequences in the syntenic regions are marked in red and indicated by red lines.

rather than as singletons after WGDs. This finding is consistent with previous studies and suggests that the retention or loss of conserved *MIRNAs* and their target genes follows the dosage balance rule. Next, the expression and functional divergence between the 2 members of the duplicate gene pairs in *MIRNA*–target interaction scenarios a, b, d, and e were investigated in *A. thaliana* (supplementary tables S17 and S18, Supplementary Material online). The target duplicates whose 2 members differentially expressed and with differential functional annotations (Gene Ontology terms) were obtained in each scenario, respectively. Using the corresponding analysis results of the duplicated gene pairs without *MIRNA* regulation as a control, we found that pairs co-regulated by duplicated *MIRNAs* (scenario a) exhibited a lower expression divergence rate (42.86% vs. 77.48%) and lower functional divergence

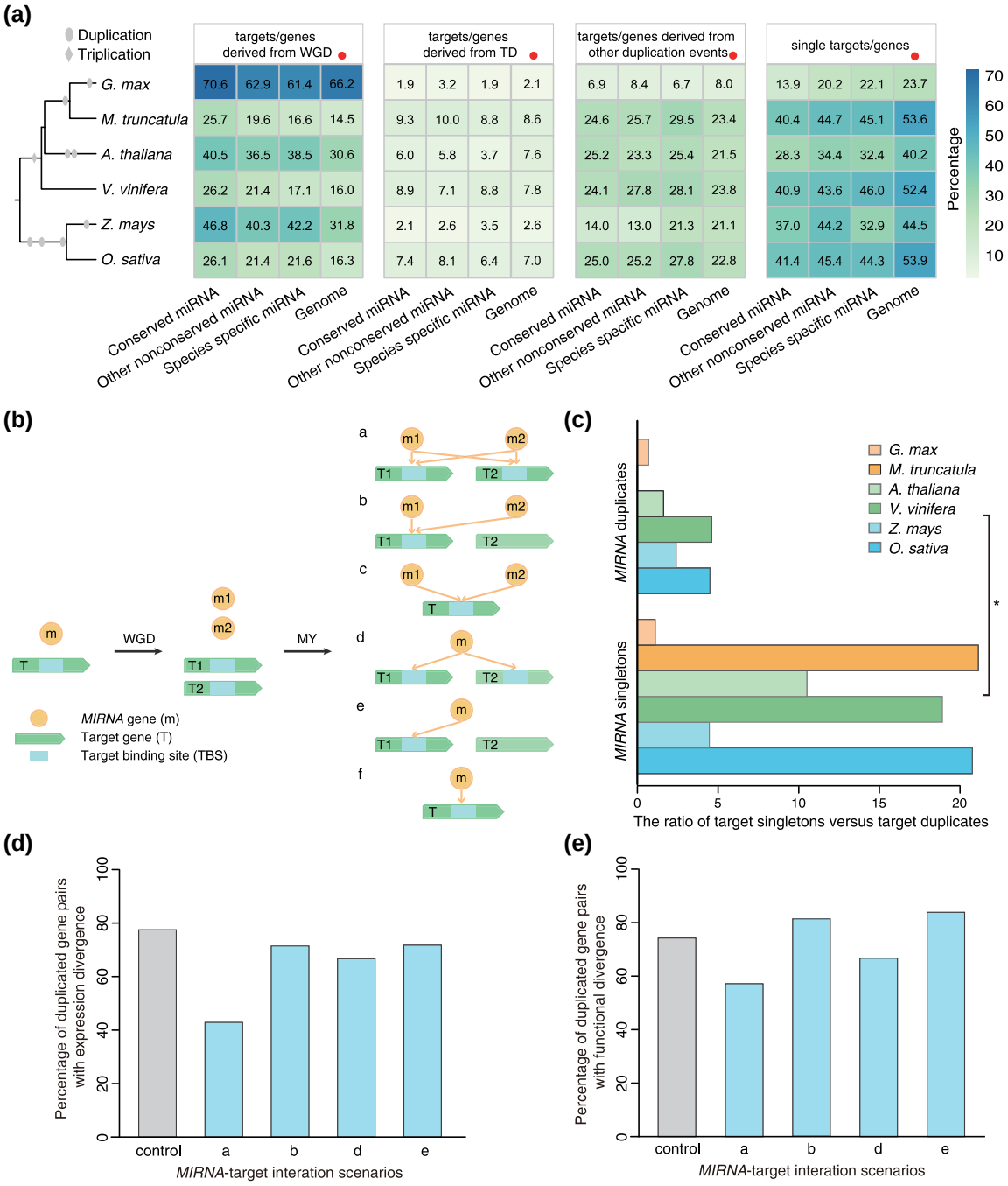

Fig. 5. The regulatory effects of miRNAs on the duplicated genes derived from WGDs. a) The proportions of duplicates resulting from different duplication mechanisms and singletons among all target genes for miRNAs with varying degrees of conservation in each species, respectively. Red dots indicate the proportion of each gene type among all protein-coding genes in each species. b) A schematic representation of different *MIRNA*–target interaction scenarios. Scenario a: both members of the *MIRNA* duplicates regulate both members of the target duplicate pairs; Scenario b: both members of the *MIRNA* duplicates regulate only one member of the target duplicate pairs. Scenario c: both members of the *MIRNA* duplicates regulate the same target singleton; Scenario d: *MIRNA* singletons regulate both members of the target duplicate pairs; Scenario e: *MIRNA* singletons regulate only one member of the target duplicate pairs; Scenario f: *MIRNA* singletons regulate target singletons. m1 and m2 represent the 2 members of miRNA duplicate pairs; T1 and T2 represent the 2 members of target duplicate pairs; m and T represent miRNA singleton and target singleton, respectively. c) Comparison of the ratio of target singletons to target duplicates between conserved *MIRNA* duplicates (the number of targets in scenario c vs. scenario a) and *MIRNA* singletons (the number of targets in scenario f vs. scenario d) in studied species. The statistical analysis was conducted by Student's *t* test, **P*-value <0.05. d) The percentage of duplicated target genes whose 2 members express divergently in each *MIRNA*–target interaction scenario for conserved *MIRNAs*. e) The percentage of duplicated target genes whose 2 members are functionally divergent in each *MIRNA*–target interaction scenario for conserved *MIRNAs*.

rate (57.14% vs. 74.23%). In contrast, the expression divergence rate of duplicated gene pairs with one member co-regulated by duplicated *MIRNAs* (scenario b) or *MIRNA* singletons (scenario e) was similar to that of pairs without *MIRNA* regulation, though the functional divergence rates in these scenarios were higher than the control (Fig. 5d and e). Therefore, these results suggest that miRNA might act as a genetic buffer against genomic shock subsequent to WGDs, and maintain the stability of expression of target genes and related regulatory networks. In addition, with the gradual loss and sequence divergence of duplicated *MIRNAs* and genes in patterns b and e, the divergence and gain/loss of *MIRNA*–target interaction would occur and result in a rewired and more complex regulatory network. Thus, it appears that miRNAs may play essential roles in stabilizing and complexing the related regulatory networks following WGDs.

Discussion

WGD, caused by polyploidization, is a widespread phenomenon in plant genomes that results in a sudden increase in the entire gene set. Like protein-coding genes, *MIRNAs* also duplicate during WGDs. As crucial regulators mainly at the posttranscriptional level, miRNAs regulate gene expression by binding to the transcripts of target genes based on sequence complementarity. Therefore, both the duplicate status (or dosage) and sequence divergence of *MIRNAs* significantly affect gene expression and the evolution of genetic networks. To understand the evolutionary patterns and consequences of conserved *MIRNAs* following WGDs, we performed genome-wide and comprehensive studies on the retention-loss, birth-death and sequence conservation-divergence of conserved *MIRNAs* in 3 pairs of species with different WGD histories. The results indicate that WGDs may play an essential role in increasing the member of conserved miRNA families and provide novel insights into the independent origination of some of their members.

Previous studies have suggested that WGD contributes to the expansion of miRNA families by analyzing miRNA evolution within specific families or species. By comprehensively investigating the retention patterns of *MIRNAs* following WGDs in 6 studied species, we also found that some *MIRNAs* were retained as duplicates, with a higher percentage in species that have undergone more rounds of WGDs. Furthermore, by identifying conserved and other *MIRNAs* in *MIRNA* duplicated pairs derived from the independent WGDs in *G. max*, *A. thaliana* and *Z. mays*, as well as in pairs derived from the shared WGDs in these 6 species, we observed that conserved *MIRNAs* account for a higher proportion of *MIRNA* duplicates compared with other *MIRNAs* in *MIRNA* pairs duplicated from whether recent or ancient WGDs. These observations reveal that conserved *MIRNAs* are preferentially retained as duplicates rather than other *MIRNAs* after WGDs. This indicates that WGDs facilitate the expansion of conserved *MIRNA* families, consistent with previous studies showing that multiple-copy *MIRNAs* are evolutionarily more conserved than single-copy *MIRNAs* (Sun et al. 2015; Zhao et al. 2015).

The conservation of miRNA families is typically defined based on their phylogenetic distribution and nucleotide sequence similarity of the mature miRNA (Chávez Montes et al. 2014; Taylor et al. 2014; You et al. 2017). However, *MIRNA* genes, like any other genes, can be gained and lost during evolution, and there is often little sequence conservation beyond the DNA encoding the mature miRNA and miRNA*. It may thus appear possible that some miRNAs within the same family may be homoplasious rather than homologous. In this study, we unambiguously identified *MIRNA* homologs from the 6 selected species using a synteny-based comparative genomics strategy. We further investigated the evolution of encoding loci of conventionally conserved miRNA families. Our results indicate that the loss of copies duplicated from pre-existing loci and the independent origination of copies in new encoding loci in specific lineages or species occurred frequently during the evolution of conserved miRNA families. Specifically, for certain members of the miR397 family (Gma-miR397c and Zma-miR397a), which are located in non-syntenic loci, we found evidence suggesting that sequence divergence of the corresponding syntenic segments may have led them to "born" independently. While we cannot entirely exclude the possibility that pseudogenization of inherited syntenic miR397 in the homologous loci of other species results in none syntenic miRNA were identified for Gma-miR397c and Zma-miR397a. We infer, based on the principle of maximum parsimony, that these 2 miRNAs likely originated independently in new loci. Thus, it is reasonable to speculate that, like non-conserved miRNA families, the conventional conserved miRNA families may also experience high birth and death rates, with their members originating independently multiple times throughout evolution, despite producing highly conserved or even identical mature miRNA sequences. Due to the limited number of species used in this study, the identification of homologous miRNA inherited from the same ancestral loci, as well as the investigation of homologous loci of non-syntenic miRNAs in other species, might be incomplete. This could lead to the inaccuracies in the birth and death analyses of conserved miRNA families. For instance, some miRNAs thought to have originated de novo may actually have syntenic miRNAs in species we did not include. Further homologous analyses, based on more representative species from major clades of land plants, could provide a better understanding of the evolution of conserved miRNA families.

It is well acknowledged that sequence divergence of protein-coding genes is prevalent following WGDs.

However, the sequence divergence of *MIRNA* duplicates after WGDs remains under-studied. To understand the sequence divergence of duplicated *MIRNAs*, we compared the mature sequences of duplicated *MIRNA* pairs in species with independent WGD events with their respective orthologs in other species. The results show that more than half of the *MIRNA* duplicates produce identical mature miRNAs, and the majority of these are also identical to the mature miRNA produced by their orthologs, suggesting high sequence conservation among duplicated *MIRNAs*. Furthermore, we compared the sequence divergence of *MIRNA* duplicates derived from different WGDs and found that *MIRNA* duplicates producing identical mature miRNAs have a higher percentage of pairs derived from recent, lineage or species specific WGD events, compared with *MIRNA* duplicates generating different miRNAs. Considering the high proportion of conserved *MIRNAs* in *MIRNA* duplicates, this result suggests that conserved *MIRNA* duplicates diverge gradually following WGD events. Like protein-coding genes, the sequence divergence of *MIRNA* duplicates may lead to sub-functionalization or neo-functionalization, contributing further to the evolution of related regulatory networks.

Duplicated gene pairs resulting from WGDs can follow several evolutionary fates, including conservation, neo-functionalization, sub-functionalization, and non-functionalization, which drive the rewiring and complexity of regulatory networks (Rensing 2014; Wendel et al. 2016). In this study, we explored the regulatory effects of conserved miRNAs on the evolutionary processes of paralogous genes following WGDs, by examining expression and functional divergence between the 2 members of duplicate gene pairs with various *MIRNA*-target interaction scenarios (a, b, d, and e) in *A. thaliana*. Our results suggest that miRNA may serve as genetic buffers against genomic shock following WGDs, leading to reduced expression and functional divergence in duplicated gene pairs in scenario a, thereby stabilizing related regulatory networks. However, as duplicated *MIRNAs* and target genes gradually lose function and diverge in sequence, *MIRNA*–target interaction may evolve. This could result in some duplicated gene copies escaping miRNAs regulation or other copies acquiring miRNA binding sites (pattern b and e), contributing to functional divergence and a more complex regulatory network. Thus, miRNAs may play a critical role in both stabilizing and complicating regulatory networks following WGDs. Further comparative analyses of the expression and functional divergence of gene duplicates with different miRNA-target interaction scenarios based on more transcriptome and experimental datasets from other species would be valuable in validating this hypothesis and enhancing our understanding of the regulatory effects of conserved miRNAs on the paralogous genes and associated biological networks following WGDs.

Materials and Methods

Genome, miRNA and Target Data Sets

The complete or draft genome sequences of the 6 species studied were obtained from Phytozome (https://phytozome.jgi.doe.gov/pz/portal.html). The miRNA and target annotation information for these species were downloaded from the PmiREN database (Guo et al. 2020). The miRNA targets annotated in this database are supported by evidence from 2 plant miRNA target prediction methods: psRNATarget (Dai et al. 2018) and RNAhybrid (Kruger and Rehmsmeier 2006). Some targets also have confirmation from PARE (Parallel Analysis of RNA Ends) sequences datasets, indicating cleavage signals in several species. Due to the high rate of false positives in prediction results, we further predicted the miRNA targets for each species using TargetFinder (Bo and Wang 2005), applying strict filtering criteria. Targets were identified as high-confidence if they met at least one of the following criteria: (i) significant cleavage signal from the PARE sequencing dataset (categories 0 or 1); (ii) cleavage signal from the PARE sequencing dataset (category 2), along with predictions from at least one of the prediction tools; (iii) weak (categories 3 or 4) or no cleavage signal from the PARE sequencing dataset, with predictions from at least 2 of the 3 prediction tools.

Homologous *MIRNAs* and Target Genes Identification by Performing Synteny Analyses

Intragenomic and intergenomic synteny analyses were performed using MCScanX (Wang et al. 2012). First, GFF (General Feature Format) files for the *MIRNAs* in each species were generated and combined with the files for all protein-coding genes within the same species. Next, we conducted an all-against-all BLASTP analysis for protein-coding genes and BLASTN for precursor sequences of *MIRNAs* using an *E*-value cutoff of 10^{-3} (Camacho et al. 2009). The top 20 BLAST matches for each query were selected, and the results for both protein-coding genes and *MIRNAs* were integrated. Finally, we utilized MCScanX to identify syntenic blocks using default parameters, which require a minimum of 5 genes to define a syntenic block and allow for a maximum of 25 gene gaps. This approach facilitated the identification of conserved genomic regions across the studied species (Wang et al. 2012).

Paralogous and orthologous pairs for *MIRNAs* and their target genes were identified based on genome synteny. The *MIRNA* pairs located within syntenic loci of intragenomic or intergenomic blocks, and generating mature miRNAs that belong to the same miRNA family, were classified as paralogous or orthologous pairs, respectively. Similarly, syntenic gene pairs, where both members are regulated by miRNAs, were referred to as paralogous or orthologous target duplicates.

Identification of *MIRNA* Duplicate Pairs Derived From Different WGDs

To identify the duplicate *MIRNA* pairs generated by various WGD events in species with independent WGDs (*A. thaliana*, *G. max,* and *Z. mays*), we analyzed the distribution of *Ks* values of orthologous gene pairs between the genomes of each pair of species. The detailed steps were as follows: (i) BLASTP Searches: we performed all-against-all BLASTP searches between the genomes with an *E*-value cutoff of 1×10^{-3} (Camacho et al. 2009), determining orthologous gene pairs by the best reciprocal matches. (ii) Protein Sequence Alignment: the protein sequences of each orthologous gene pair were aligned using MUSCLE (Edgar 2004), and the resulting protein alignments guided coding sequence alignments using PAL2NAL (Suyama et al. 2006). (iii) *Ks* Calculation: the *Ks* value for each orthologous gene pair was calculated using the Nei-Gojobori algorithm (Nei and Gojobori 1986) implemented in the yn00 program of the PAML package (Yang 2007). (iv) Filtering: gene pairs with *Ks* values greater than 3.0 were excluded due to saturated substitutions at synonymous sites. The *Ks* distribution of the remaining gene pairs was then constructed. (v) Peak Identification: the peak of the *Ks* distribution was identified and the *Ks* value at this peak was used as a cutoff in subsequent analyses. Next, we calculated the *Ks* value of each gene pair in the syntenic blocks of *A. thaliana*, *G. max,* and *Z. mays* following the previously mentioned steps. We then computed the median *Ks* values of all syntenic gene pairs within each syntenic block. Intragenomic syntenic *MIRNA* pairs in *A. thaliana*, *G. max,* and *Z. mays* with median *Ks* value less than their respective cutoff values were considered to be derived from recent, lineage- or species-specific WGDs. Conversely, syntenic *MIRNA* duplicates with median *Ks* values greater than their respective cutoff values were considered to be derived from shared WGDs of the pairwise species.

Cluster Analysis of *MIRNA*

To comprehensively and accurately identify the homologous *MIRNA* cluster, we further conducted synteny analyses using MCScanX with the maximum gap between the anchor genes to 50 (Wang et al. 2012). Following this, the homologous *MIRNAs*, which are inherited from the same ancestral encoding locus were grouped into the same cluster using a custom Python script. For *MIRNA* within the same species or branch (core eudicots or monocots), intergenomic and intragenomic syntenic *MIRNA* pairs with overlapping members were grouped into the same cluster, based on the synteny results with a maximum gap of 25 genes. For *MIRNAs* of eudicot and monocot species, the paralogous and orthologous *MIRNA* pairs with overlapping members were group together if *MIRNA* from one lineage showed syntenic relationships with at least 2 species from the other lineage, based on the synteny results with a maximum gap of 50 genes.

Identification and Classification of Target Genes Derived From Different Duplication Events

To identify target duplicates derived from various duplication events, we conducted the following analyses for each species: (i) detection of duplicate gene pairs: all duplicate gene pairs were identified using best reciprocal BLASTP alignments (Camacho et al. 2009). (ii) Classification of duplicate gene pairs: these duplicate gene pairs were classified into 3 categories: WGD-derived pairs—gene pairs with evidence of genomic synteny; tandem duplicated pairs—gene pairs located within 5 genetic loci of each other; other duplicate pairs—the remaining duplicate gene pairs. (iii) Target identification: miRNA targets in each category were further identified and considered as the target duplicates derived from the corresponding duplication events.

Expression Divergence and Functional Relevance Analyses of the Duplicated Gene Pairs With Different miRNA–Target Interaction Patterns

To better understand the significant roles of conserved miRNA in the regulatory network following WGDs, we estimated the expression divergence and functional relevance of paralogous gene pairs with different miRNA-target interaction patterns in *A. thaliana*. The steps were as follows: (i) expression data collection: SRA data files for transcriptomes of 4 tissues (flower, leaf, siliques, and seedling) of *A. thaliana* were downloaded from the Gene Expression Omnibus (GEO accessions GSE34318, GSE61545, and GSE49950) using "prefetch" command from sratoolkit.3.0.0 and were further converted into FASTQ format with the "fastq-dump" command. (ii) Transcriptomic data analyses: these raw RNA sequencing reads were preprocessed using Trimmomatic (Bolger et al. 2014) to remove adaptor sequences and low-quality reads. The clean reads were then mapped to the reference genome using HISAT2 with default parameters. The expression abundance values were calculated using Stringtie (Pertea et al. 2016), and the abundance values from the 3 biological replicates of each sample were averaged to obtain the expression level of each gene. (iii) Retrieval of duplicated gene pairs: we retrieved duplicated gene pairs with miRNA-target interaction pattern a, b, d, and e for conserved miRNAs. (iv) Expression divergence analysis: for each duplicated gene pair, the fold change of RPKM values between the 2 members in each tissue was calculated. Pairs with an absolute value of $\log_2$ fold change ≥ 1 in at least one tissue were considered to be differentially expressed. (v) Functional relevance investigation: to assess functional relevance, the functional annotation of the 2 members in each duplicated gene pair was compared based on the gene ontology (GO) annotations of *A. thaliana* download

from The *Arabidopsis* Information Resource (https://www.arabidopsis.org/) database. The 4 categories of duplicated gene pairs with different GO terms were identified, and these pairs were considered functionally divergent. (vi) Control analysis: duplicated gene pairs without miRNA regulation were also identified and their corresponding analysis results were used as controls in subsequent comparative analyses.

Supplementary Material

Supplementary material is available at *Genome Biology and Evolution* online.

Acknowledgments

We thank Prof. Hongzhi Kong and Xiaozeng Yang for helpful discussions on related analysis.

Author Contributions

Y.J. conceived and designed the research. L.Q. performed the research. P.X. identified the miRNA cluster by custom Python script. L.Q. and Y.J. wrote the manuscript. All authors have read and approved the manuscript.

Funding

This study was supported by CAS Youth Interdisciplinary Team (JCTD-2022-06) and National Natural Science Foundation of China (32200182).

Conflict of Interest

The authors declare that there are no conflicts of interest associated with this study.

Data Availability

All data generated or analyzed during this study are included in this published article and its supplementary information files.

Literature Cited

Abrouk M, Zhang R, Murat F, Li A, Pont C, Mao L, Salse J. Grass microRNA gene paleohistory unveils new insights into gene dosage balance in subgenome partitioning after whole-genome duplication. Plant Cell. 2012:24(5):1776–1792. https://doi.org/10.1105/tpc.112.095752.

Allen E, Xie Z, Gustafson AM, Sung G-H, Spatafora JW, Carrington JC. Evolution of microRNA genes by inverted duplication of target gene sequences in *Arabidopsis thaliana*. Nat Genet. 2004:36(12):1282–1290. https://doi.org/10.1038/ng1478.

Axtell MJ. Classification and comparison of small RNAs from plants. Annu Rev Plant Biol. 2013a:64(1):137–159. https://doi.org/10.1146/annurev-arplant-050312-120043.

Axtell MJ. ShortStack: comprehensive annotation and quantification of small RNA genes. RNA. 2013b:19(6):740–751. https://doi.org/10.1261/rna.035279.112.

Axtell MJ, Bowman JL. Evolution of plant microRNAs and their targets. Trends Plant Sci. 2008:13(7):343–349. https://doi.org/10.1016/j.tplants.2008.03.009.

Axtell MJ, Meyers BC. Revisiting criteria for plant microRNA annotation in the era of big data. Plant Cell. 2018:30(2):272–284. https://doi.org/10.1105/tpc.17.00851.

Baldrich P, Beric A, Meyers BC. Despacito: the slow evolutionary changes in plant microRNAs. Curr Opin Plant Biol. 2018:42:16–22. https://doi.org/10.1016/j.pbi.2018.01.007.

Bo X, Wang S. TargetFinder: a software for antisense oligonucleotide target site selection based on MAST and secondary structures of target mRNA. Bioinformatics. 2005:21(8):1401–1402. https://doi.org/10.1093/bioinformatics/bti211.

Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for illumina sequence data. Bioinformatics. 2014:30(15):2114–2120. https://doi.org/10.1093/bioinformatics/btu170.

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. BLAST+: architecture and applications. BMC Bioinformatics. 2009:10(1):421. https://doi.org/10.1186/1471-2105-10-421.

Carthew RW, Sontheimer EJ. Origins and mechanisms of miRNAs and siRNAs. Cell. 2009:136(4):642–655. https://doi.org/10.1016/j.cell.2009.01.035.

Chávez Montes RA, Rosas-Cárdenas dFF, De Paoli E, Accerbi M, Rymarquis LA, Mahalingam G, Marsch-Martínez N, Meyers BC, Green PJ, de Folter S. Sample sequencing of vascular plants demonstrates widespread conservation and divergence of microRNAs. Nat Commun. 2014:5(1):3722. https://doi.org/10.1038/ncomms4722.

Chen C, Li J, Feng J, Liu B, Feng L, Yu X, Li G, Zhai J, Meyers BC, Xia R. sRNAanno-a database repository of uniformly annotated small RNAs in plants. Hortic Res. 2021:8(1):45. https://doi.org/10.1038/s41438-021-00480-8.

Chen C, Zeng Z, Liu Z, Xia R. Small RNAs, emerging regulators critical for the development of horticultural traits. Hortic Res. 2018:5(1):63. https://doi.org/10.1038/s41438-018-0072-8.

Cui J, You C, Chen X. The evolution of microRNAs in plants. Curr Opin Plant Biol. 2017:35:61–67. https://doi.org/10.1016/j.pbi.2016.11.006.

Cuperus JT, Fahlgren N, Carrington JC. Evolution and functional diversification of *MIRNA* genes. Plant Cell. 2011:23(2):431–442. https://doi.org/10.1105/tpc.110.082784.

D'Ario M, Griffiths-Jones S, Kim M. Small RNAs: big impact on plant development. Trends Plant Sci. 2017:22(12):1056–1068. https://doi.org/10.1016/j.tplants.2017.09.009.

Dai X, Zhuang Z, Zhao PX. psRNATarget: a plant small RNA target analysis server (2017 release). Nucleic Acids Res. 2018:46(W1):W49–W54. https://doi.org/10.1093/nar/gky316.

De Smet R, Van de Peer Y. Redundancy and rewiring of genetic networks following genome-wide duplication events. Curr Opin Plant Biol. 2012:15(2):168–176. https://doi.org/10.1016/j.pbi.2012.01.003.

Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004:32(5):1792–1797. https://doi.org/10.1093/nar/gkh340.

Fahlgren N, Howell MD, Kasschau KD, Chapman EJ, Sullivan CM, Cumbie JS, Givan SA, Law TF, Grant SR, Dangl JL, et al. High-throughput sequencing of *Arabidopsis* microRNAs: evidence for frequent birth and death of *MIRNA* genes. PLoS One. 2007:2(2):e219. https://doi.org/10.1371/journal.pone.0000219.

Felippes FF, Schneeberger K, Dezulian T, Huson DH, Weigel D. Evolution of *Arabidopsis thaliana* microRNAs from random sequences. RNA. 2008:14(12):2455–2459. https://doi.org/10.1261/rna.1149408.

Fox DT, Soltis DE, Soltis PS, Ashman TL, Van de Peer Y. Polyploidy: a biological force from cells to ecosystems. Trends Cell Biol. 2020:30(9):688–694. https://doi.org/10.1016/j.tcb.2020.06.006.

Guo Z, Kuang Z, Tao Y, Wang H. Miniature inverted-repeat transposable elements drive rapid microRNA diversification in angiosperms. Mol Biol Evol. 2022:39(11):msac224. https://doi.org/10.1093/molbev/msac224.

Guo Z, Kuang Z, Wang Y, Zhao Y, Tao Y, Cheng C, Yang J, Lu X, Hao C, Wang T, et al. PmiREN: a comprehensive encyclopedia of plant miRNAs. Nucleic Acids Res. 2020:48(D1):D1114–D1121. https://doi.org/10.1093/nar/gkz894.

He H, Liang G, Li Y, Wang F, Yu D. Two young microRNAs originating from target duplication mediate nitrogen starvation adaptation via regulation of glucosinolate synthesis in *Arabidopsis thaliana*. Plant Physiol. 2014:164(2):853–865. https://doi.org/10.1104/pp.113.228635.

Huang HY, Lin Y-C-D, Cui S, Huang Y, Tang Y, Xu J, Bao J, Li Y, Wen J, Zuo H, et al. miRTarBase update 2022: an informative resource for experimentally validated miRNA-target interactions. Nucleic Acids Res. 2022:50(D1):D222–D230. https://doi.org/10.1093/nar/gkab1079.

Jiao Y. Double the genome, double the fun: genome duplications in angiosperms. Mol Plant. 2018:11(3):357–358. https://doi.org/10.1016/j.molp.2018.02.009.

Jiao Y, Leebens-Mack J, Ayyampalayam S, Bowers JE, McKain MR, McNeal J, Rolf M, Ruzicka DR, Wafula E, Wickett NJ, et al. A genome triplication associated with early diversification of the core eudicots. Genome Biol. 2012:13(1):R3. https://doi.org/10.1186/gb-2012-13-1-r3.

Jiao Y, Li J, Tang H, Paterson AH. Integrated syntenic and phylogenomic analyses reveal an ancient genome duplication in monocots. Plant Cell. 2014:26(7):2792–2802. https://doi.org/10.1105/tpc.114.127597.

Jiao Y, Wickett NJ, Ayyampalayam S, Chanderbali AS, Landherr L, Ralph PE, Tomsho LP, Hu Y, Liang H, Soltis PS, et al. Ancestral polyploidy in seed plants and angiosperms. Nature. 2011:473(7345):97–100. https://doi.org/10.1038/nature09916.

Kozomara A, Birgaoanu M, Griffiths-Jones S. miRBase: from microRNA sequences to function. Nucleic Acids Res. 2019:47(D1):D155–D162. https://doi.org/10.1093/nar/gky1141.

Kruger J, Rehmsmeier M. RNAhybrid: microRNA target prediction easy, fast and flexible. Nucleic Acids Res. 2006:34(Web Server):W451–W454. https://doi.org/10.1093/nar/gkl243.

Kuang Z, Wang Y, Li L, Yang X. miRDeep-P2: accurate and fast analysis of the microRNA transcriptome in plants. Bioinformatics. 2019:35(14):2521–2522. https://doi.org/10.1093/bioinformatics/bty972.

Li A, Mao L. Evolution of plant microRNA gene families. Cell Res. 2007:17(3):212–218. https://doi.org/10.1038/sj.cr.7310113.

Li Y, Li CQ, Xia J, Jin YX. Domestication of transposable elements into microRNA genes in plants. PLoS One. 2011:6(5):e19212. https://doi.org/10.1371/journal.pone.0019212.

Lohaus R, Van de Peer Y. Of dups and dinos: evolution at the K/pg boundary. Curr Opin Plant Biol. 2016:30:62–69. https://doi.org/10.1016/j.pbi.2016.01.006.

Lunardon A, Johnson NR, Hagerott E, Phifer T, Polydore S, Coruh C, Axtell MJ. Integrated annotations and analyses of small RNA-producing loci from 47 diverse plants. Genome Res. 2020:30(3):497–513. https://doi.org/10.1101/gr.256750.119.

Maher C, Stein L, Ware D. Evolution of *Arabidopsis* microRNA families through duplication events. Genome Res. 2006:16(4):510–519. https://doi.org/10.1101/gr.4680506.

Meyers BC, Axtell MJ. MicroRNAs in plants: key findings from the early years. Plant Cell. 2019:31(6):1206–1207. https://doi.org/10.1105/tpc.19.00310.

Meyers BC, Axtell MJ, Bartel B, Bartel DP, Baulcombe D, Bowman JL, Cao X, Carrington JC, Chen X, Green PJ, et al. Criteria for annotation of plant microRNAs. Plant Cell. 2008:20(12):3186–3190. https://doi.org/10.1105/tpc.108.064311.

Moore RC, Purugganan MD. The evolutionary dynamics of plant duplicate genes. Curr Opin Plant Biol. 2005:8(2):122–128. https://doi.org/10.1016/j.pbi.2004.12.001.

Nei M, Gojobori T. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol Biol Evol. 1986:3(5):418–426. https://doi.org/10.1093/oxfordjournals.molbev.a040410.

Panchy N, Lehti-Shiu M, Shiu SH. Evolution of dene duplication in plants. Plant Physiol. 2016:171(4):2294–2316. https://doi.org/10.1104/pp.16.00523.

Pasquinelli AE. MicroRNAs and their targets: recognition, regulation and an emerging reciprocal relationship. Nat Rev Genet. 2012:13(4):271–282. https://doi.org/10.1038/nrg3162.

Pertea M, Kim D, Pertea GM, Leek JT, Salzberg SL. Transcript-level expression analysis of RNA-Seq experiments with HISAT, StringTie and Ballgown. Nat Protoc. 2016:11(9):1650–1667. https://doi.org/10.1038/nprot.2016.095.

Piriyapongsa J, Jordan IK. Dual coding of siRNAs and miRNAs by plant transposable elements. RNA. 2008:14(5):814–821. https://doi.org/10.1261/rna.916708.

Rensing SA. Gene duplication as a driver of plant morphogenetic evolution. Curr Opin Plant Biol. 2014:17:43–48. https://doi.org/10.1016/j.pbi.2013.11.002.

Roulin A, Auer PL, Libault M, Schlueter J, Farmer A, May G, Stacey G, Doerge RW, Jackson SA. The fate of duplicated genes in a polyploid plant genome. Plant J. 2013:73(1):143–153. https://doi.org/10.1111/tpj.12026.

Rubio-Somoza I, Weigel D. MicroRNA networks and developmental plasticity in plants. Trends Plant Sci. 2011:16(5):258–264. https://doi.org/10.1016/j.tplants.2011.03.001.

Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, Nelson W, Hyten DL, Song Q, Thelen JJ, Cheng J, et al. Genome sequence of the palaeopolyploid soybean. Nature. 2010:463(7278):178–183. https://doi.org/10.1038/nature08670.

Soltis PS, Soltis DE. Ancient WGD events as drivers of key innovations in angiosperms. Curr Opin Plant Biol. 2016:30:159–165. https://doi.org/10.1016/j.pbi.2016.03.015.

Song X, Li Y, Cao X, Qi Y. MicroRNAs and their regulatory roles in plant–environment. Annu Rev Plant Biol. 2019:70(1):489–525. https://doi.org/10.1146/annurev-arplant-050718-100334.

Sun C, Wu J, Liang J, Schnable JC, Yang W, Cheng F, Wang X. Impacts of whole-genome triplication on *MIRNA* evolution in *Brassica rapa*. Genome Biol Evol. 2015:7(11):3085–3096. https://doi.org/10.1093/gbe/evv206.

Suyama M, Torrents D, Bork P. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. Nucleic Acids Res. 2006:34(Web Server):W609–W612. https://doi.org/10.1093/nar/gkl315.

Taylor RS, Tarver JE, Hiscock SJ, Donoghue PC. Evolutionary history of plant microRNAs. Trends Plant Sci. 2014:19(3):175–182. https://doi.org/10.1016/j.tplants.2013.11.008.

Van de Peer Y, Mizrachi E, Marchal K. The evolutionary significance of polyploidy. Nat Rev Genet. 2017:18(7):411–424. https://doi.org/10.1038/nrg.2017.26.

Voinnet O. Origin, biogenesis, and activity of plant microRNAs. Cell. 2009:136(4):669–687. https://doi.org/10.1016/j.cell.2009.01.046.

Wang X, Ma X, Yan G, Hua L, Liu H, Huang W, Liang Z, Chao Q, Hibberd JM, Jiao Y, et al. Gene duplications facilitate C4-CAM compatibility in common purslane. Plant Physiol. 2023:193(4):2622–2639. https://doi.org/10.1093/plphys/kiad451.

Wang X, Wang H, Wang J, Sun R, Wu J, Liu S, Bai Y, Mun J-H, Bancroft I, Cheng F, et al. The genome of the mesopolyploid crop species *Brassica rapa*. Nat Genet. 2011:43(10):1035–1039. https://doi.org/10.1038/ng.919.

Wang X, Yan X, Hu Y, Qin L, Wang D, Jia J, Jiao Y. A recent burst of gene duplications in triticeae. Plant Commun. 2021:3(2):100268. https://doi.org/10.1016/j.xplc.2021.100268.

Wang Y, Tang H, Debarry JD, Tan X, Li J, Wang X, Lee T-H, Jin H, Marler B, Guo H, et al. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. Nucleic Acids Res. 2012:40(7):e49. https://doi.org/10.1093/nar/gkr1293.

Wendel JF, Jackson SA, Meyers BC, Wing RA. Evolution of plant genome architecture. Genome Biol. 2016:17(1):37. https://doi.org/10.1186/s13059-016-0908-1.

Wu HJ, Ma YK, Chen T, Wang M, Wang XJ. Psrobot: a web-based plant small RNA meta-analysis toolbox. Nucleic Acids Res. 2012:40(W1):W22–W28. https://doi.org/10.1093/nar/gks554.

Wu S, Han B, Jiao Y. Genetic contribution of paleopolyploidy to adaptive evolution in angiosperms. Mol Plant. 2020:13(1):59–71. https://doi.org/10.1016/j.molp.2019.10.012.

Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol. 2007:24(8):1586–1591. https://doi.org/10.1093/molbev/msm088.

You C, Cui J, Wang H, Qi X, Kuo L-Y, Ma H, Gao L, Mo B, Chen X. Conservation and divergence of small RNA pathways and microRNAs in land plants. Genome Biol. 2017:18(1):158. https://doi.org/10.1186/s13059-017-1291-2.

Yu Y, Zhang Y, Chen X, Chen Y. Plant noncoding RNAs: hidden players in development and stress responses. Annu Rev Cell Dev Biol. 2019:35(1):407–431. https://doi.org/10.1146/annurev-cellbio-100818-125218.

Zhan J, Meyers BC. Plant small RNAs: their biogenesis, regulatory roles, and functions. Annu Rev Plant Biol. 2023:74(1):21–51. https://doi.org/10.1146/annurev-arplant-070122-035226.

Zhao M, Meyers BC, Cai C, Xu W, Ma J. Evolutionary patterns and coevolutionary consequences of *MIRNA* genes and microRNA targets triggered by multiple mechanisms of genomic duplications in soybean. Plant Cell. 2015:27(3):546–562. https://doi.org/10.1105/tpc.15.00048.

Associate editor: Selene Fernandez-Valverde