


RESEARCH PAPER



De novo origination of *MIRNAs* through generation of short inverted repeats in target genes

Shanfa Lu 

Institute of Medicinal Plant Development, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing, China

ABSTRACT

MIRNA (*MIR*) gene origin and early evolutionary processes, such as hairpin precursor sequence origination, promoter activity acquisition and the sequence of these two processes, are fundamental and fascinating subjects. Three models, including inverted gene duplication, spontaneous evolution and transposon transposition, have been proposed for *de novo* origination of hairpin precursor sequence. However, these models still open to discussion. In addition, *de novo* origination of *MIR* gene promoters has not been well investigated. Here, I systematically investigated the origin of evolutionarily young polyphenol oxidase gene (*PPO*)-targeting *MIRs*, including *MIR1444*, *MIR058* and *MIR12112*, and a genomic region termed *AasPPO-as-hp*, which contained a hairpin-forming sequence. I found that *MIR058* precursors and the hairpin-forming sequence of *AasPPO-as-hp* originated in an ancient *PPO* gene through forming short inverted repeats. Palindromic-like sequences and imperfect inverted repeats in the ancient *PPO* gene contributed to initiate the generation of short inverted repeats probably by causing errors during DNA duplication. Analysis of *MIR058* and *AasPPO-as-hp* promoters showed that they originated in the 3'-flanking region of the ancient *PPO* gene. Promoter activities were gained by insertion of a CAAT-box and multiple-copper-response element (CuRE)-containing miniature inverted-repeat transposable element (MITE) in the upstream of AT-rich TATA-box-like sequence. Gain of promoter activities occurred before hairpin-forming sequence origination. Sequence comparison of *MIR1444*, *MIR058* and *MIR12112* promoters showed frequent birth and death of CuREs, indicating copper could be vital for the origination and evolution of *PPO*-targeting *MIRs*. Based on the evidence obtained, a novel model for plant *MIR* origination and evolution is proposed.

ARTICLE HISTORY

Received 6 November 2018
Revised 6 March 2019
Accepted 6 March 2019

KEYWORDS

Gene origination; gene evolution; *MIR1444*; *MIR058*; *MIR12112*; polyphenol oxidase

Introduction

MicroRNAs (miRNAs) are a class of small non-coding RNAs of about 21 nucleotides in length. They are generated from primary-*MIRNAs* (*pri-MIRs*) transcribed from *MIR* loci [1]. *Pri-MIR* typically contains an imperfect hairpin structure, which is processed to a *MIRNA* precursor (*pre-MIR*). Cleavage of the *pre-MIR* on the stem portion generates a miRNA:miRNA* duplex, which comprises a mature miRNA on a side and a miRNA* on the other side. The duplex is then unwound to release a single-stranded mature miRNA [1]. Through direct cleavage of target transcripts, miRNAs play significant regulatory roles in plant development and stress responses. *De novo* origination and subsequent evolution of *MIRNA* genes (*MIRs*) is one of the most fundamental and fascinating subjects of plant biology. Origination of *MIRs* includes two key processes: the origination of hairpin precursor sequence and the acquisition of promoter activity. For *de novo* origination of *MIR* gene promoters, it has not been well investigated. For *de novo* origination of plant *pri-MIRs*, there are three models, including inverted gene duplication, spontaneous evolution and transposon transposition (Figure 1) [2–7].

The inverted gene duplication model assumed that plant *pre-MIRs* were originated by direct inverted duplication of a target gene, integration of a pseudogene-like sequence after reverse transcription, or juxtaposition of two closely related sequences from different members of a gene family (Figure 1A) [2]. This model was originally proposed based on the observation that *Arabidopsis thaliana* *MIR161* and *MIR163* precursor hairpin sequences showed extensive similarity to their target genes [2]. Later on, it was supported by the findings that some other *MIR* precursor sequences, such as *MIR482*, *MIR824*, *MIR846* and *MIR859*, contained at least one arm with similarity or complementarity to target genes [3–5]. The spontaneous evolution model assumed that the precursors of evolutionarily young *MIRs* were originated from high density of small-to-medium sized fold-back sequences scattered throughout the plant genome (Figure 1B) [1,6]. Evidence supporting this model is that some evolutionarily young *A. thaliana* *MIRs*, including *MIR774–MIR776*, *MIR779*, *MIR823*, *MIR830*, *MIR858*, *MIR864*, *MIR865* and *MIR870*, have no similarity to other regions of the *A. thaliana* genome [6]. The transposon transposition model proposed that the precursors of some plant *MIRs* were derived from transposable elements (Figure 1C) [7]. This model relied on the observation that a subset of *A. thaliana* and rice *MIR* candidates was collocated

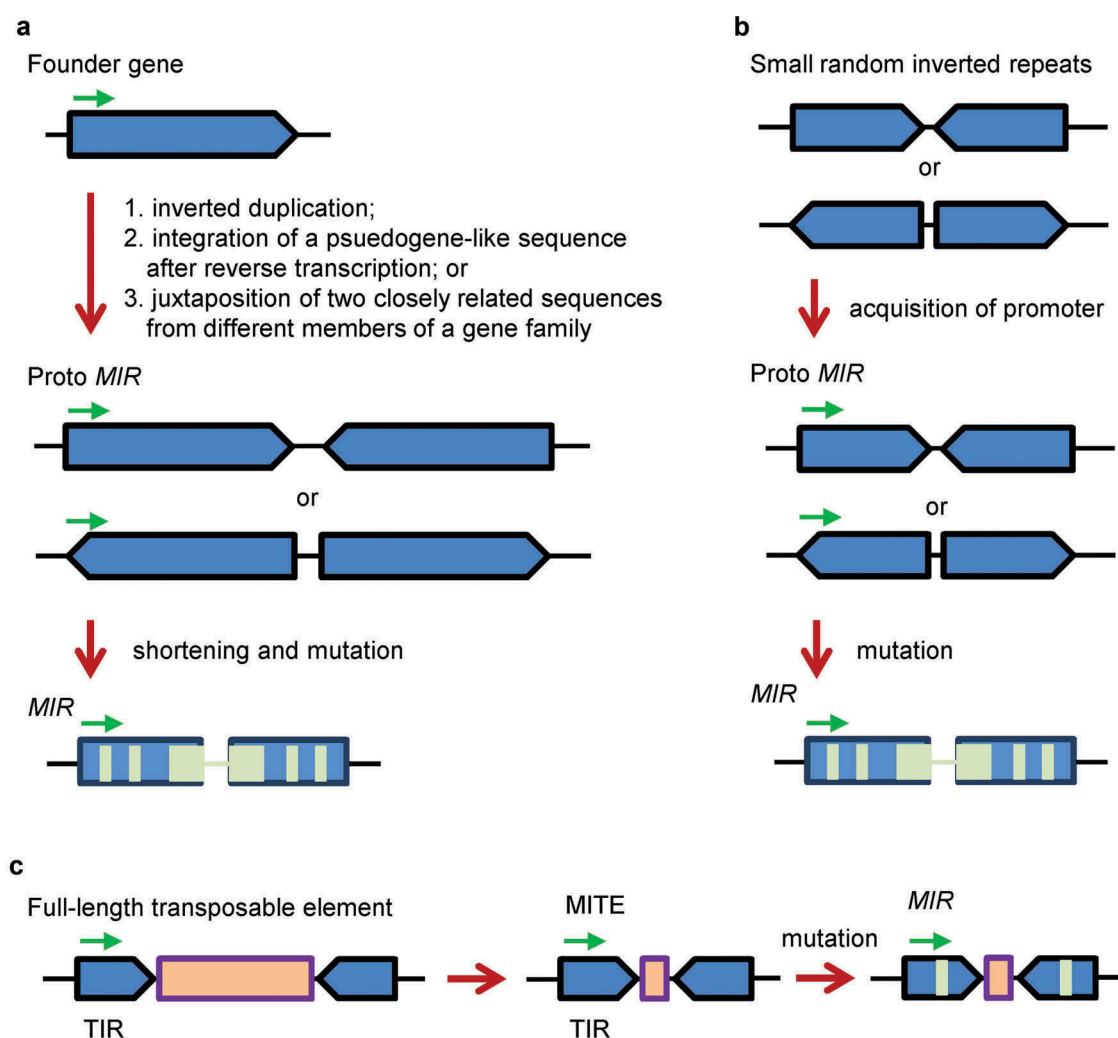


Figure 1. The inverted gene duplication, spontaneous evolution and transposon transposition models for *de novo* origination and evolution of plant *MIRs*. (A) The inverted gene duplication model [2]. Direct inverted duplication of a founder gene, integration of a pseudogene-like sequence after reverse transcription or juxtaposition of two closely related sequences from different members of a gene family resulted in the emergence of a proto *MIR*. Shortening and mutations (light green) of the proto *MIR* led to the generation of *MIRs*. (B) The spontaneous evolution model [6]. Small random inverted repeats scattered along plant genomes acquired a promoter, resulting in the emergence of a proto *MIR*. Mutations (light green) of the proto *MIR* led to the generation of *MIRs*. (C) The transposon transposition model [7]. A full-length transposable element with terminal inverted repeats (TIRs) and a long open reading frame (light brown) gave rise to a miniature inverted-repeat transposable element (MITE) with TIRs and a short open reading frame. Mutations (light green) of the MITE led to the generation of *MIRs*. Green arrows indicate transcription direction.

with miniature inverted-repeat transposable elements (MITEs) [7]. However, it could be debatable for these candidates to be bona fide *MIRs* [1]. Indeed, many of them, such as *A. thaliana* and rice *MIR416*, rice *MIR420*, *MIR445a*, *MIR806b*, *MIR806g*, *MIR807b*, *MIR807c*, *MIR809h*, *MIR811a-c*, *MIR813*, *MIR819a*, *MIR819d*, *MIR819g*, *MIR819h* and *MIR819f*, have been questioned or removed from miRBase (release 22, <http://www.mirbase.org/index.shtml>). In addition, a pre-existing *MIR* precursor might evolve into a novel one through co-evolution or perhaps punctuated jumps in sequence diversity [8]. For instance, *MIR390* could evolve into *MIR4376*, which further evolved into *MIR7122* [8].

Although three models have been proposed for *de novo* origination of plant *pre-MIRs*, detailed information for early evolutionary processes and direct evidence to support the models are still lacking. In addition, there are many issues remained to be addressed, such as the origin and early evolutionary processes of *MIR* gene promoters, the sequence of

MIR origination and promoter activity gain, and the evolutionary force driving the origination and evolution of *MIRs*. One of the difficulties to precisely elucidate the origin and early evolutionary processes of plant *MIRs* is that the majority of them have a long evolutionary history. Much sequence information required for the birth of these *MIRs* has largely been lost as time goes on. Therefore, identification of very recently evolved *MIRs* is significant for elucidation of plant *MIR* origination, since the sequence features acquired during origination may be well-preserved in these *MIRs*.

Recently, three polyphenol oxidase (*PPO*) gene-targeting and lineage-specific young *MIR* families, termed *MIR1444*, *MIR058* and *MIR12112*, have been reported in *Populus*, *Vitis* and *Salvia*, respectively [9–14]. *MIR1444* is a Salicaceae-specific *MIR* widely existing in Salicaceae, including *Populus*, *Salix* and *Idesia* [9–14]. Mature miR1444 targets a subset of *PPOs* in a region encoding the conserved CuB domain [12]. Precursor sequence of *MIR1444* exhibits extensive sequence

similarity to the *PPO* targets and was proposed to originate from an ancient *PPO* gene before the Salicoid whole-genome duplication event, which happened 60 million years ago (Ma) before the divergence of *Populus*, *Salix* and *Idesia* [12,15]. No sequence similarity was observed for the other parts of *MIR1444* and *PPO* genes. It indicates that much sequence information for tracing the early evolutionary processes of *MIR1444* has been lost during its evolution for more than 60 million years. *MIR058* is the secondly reported *PPO*-targeting *MIR* [13]. Its mature sequence was identified through the analysis of high-throughput small RNA sequencing data from grapevines [16]. Cleavage of miR058 on *PPO* transcripts in a region encoding the thylakoid transfer domain has been validated through degradome sequencing and RLM-RACE analysis [13,14,16]. The expression of miR058 and *PPO* exhibited negative correlation in various grapevine tissues [13]. These results suggest that *MIR058* is a bona fide *MIR*, although it has not been deposited into miRBase [17]. The third *PPO*-targeting *MIR*, termed *MIR12112* [14], was identified by analysis of high-throughput small RNA sequence and the whole genome sequence of *Salvia miltiorrhiza*, a well-known material of traditional Chinese medicine widely used for cardiovascular and cerebrovascular disease treatment and an emerging model system for genomic and genetic studies of medicinal plants [18–21]. This miRNA targets to 15 of the 19 identified *SmPPOs* in a region encoding the conserved KFDV domain [14]. Except for *Vitis vinifera* and *S. miltiorrhiza*, *MIR058* and *MIR12112* genes in other plant species have not been analysed.

In addition to evolutionarily young, there are multiple copies of copper (Cu)-response elements (CuREs) existing in the promoter of *Populus trichocarpa* *MIR1444* genes [11]. These sequences could be important during the origination and evolution of *MIRs*. CuRE contains a core sequence of GTAC and was firstly identified in the promoters of cytochrome *C6* (*CYC6*) and coprogen oxidase (*CPX1*) genes in the green alga *Chlamydomonas reinhardtii* [22,23]. Nowadays, people have learned that CuRE is a significant *cis*-element widely existing in promoters of Cu-responsive protein genes and plant *MIRs*, such as *MIR397*, *MIR398*, *MIR408*, and *MIR1444* [11–24]. All of these *MIRs* regulate the expression of genes encoding Cu-binding proteins, such as laccases (LACs), Cu-Zn superoxide dismutases (CSDs), plastocyanin-like proteins (PCLs), and PPOs. The expression of these *MIRs* was negatively regulated by Cu, an essential mineral required for the healthy growth and development of plants [11,25–27]. Cu-associated suppression of *MIRs* was in company with the up-regulation of their target genes [11]. *SQUAMOSA* promoter binding protein-like 7 (*SPL7*) could bind to this element to function as a central regulator for Cu homeostasis in *A. thaliana* [28]. In *P. trichocarpa*, PtSPL3 and PtSPL4 could be the CuRE-binding proteins controlling Cu-responsive gene expression [11], whereas in green alga, the CuRE-binding protein is known as COPPER RESPONSE REGULATOR 1 (*CRR1*) [29]. Through the analysis of *CRR1*, a model was proposed for Cu sensing in green alga [29]. In this model, the SPB domain of *CRR1* binds to CuRE and activates gene expression when Cu is deficient, whereas when Cu is sufficient, Cu binds to the SPB domain, resulting in *CRR1* structure transformation and

CuRE-binding capability loss [28–30]. These results suggest the significance of CuREs in plant response to Cu availability. However, there is no information about the origin and variation of CuREs in gene promoters. In addition, CuREs in the promoters of *MIR058* and *MIR12112* have not been investigated.

With long-term interests in plant miRNAs [9–12,14,31–33], particularly those targeting Cu-binding protein genes, I performed a systematic analysis of *PPO*-targeting and lineage-specific young *MIRs* to address various fundamental questions of their origination and evolution, such as how did the hairpin sequence originate, how did the promoter activity gain, what was the sequence of *pre-MIR* origination and promoter activity gain, and what could be the role of Cu in the origination and evolution of *PPO*-targeting *MIRs*? Integrative sequence analysis of genome and transcriptome data allowed me to identify a total of ten *MIR058* precursors (*pre-MIR058s*), 79 *MIR12112* precursors (*pre-MIR12112s*), and a genomic region termed *AasPPO-as-hp*, which contained a hairpin-forming sequence. Sequence comparison clearly showed that *pre-MIR058s* and the hairpin-forming sequence of *AasPPO-as-hp* originated in an ancient *PPO* gene through generation of short inverted repeats. Promoter sequence analysis showed that promoter activities were gained by insertion of a MITE sequence in the 3'-flanking region of the ancient *PPO* genes before the origination of hairpin-forming sequences. Analysis of CuREs showed the existence of multiple CuREs in the promoters of *MIR12112*, *MIR058* and *MIR1444*, with the number of CuREs varying significantly. It suggests frequent birth and death of CuREs in promoters of these *MIRs* and indicates the significance of Cu in the origination and evolution of *PPO*-targeting *MIRs*. The results reveal novel mechanistic insights into *MIR* origination and provide first-hand information for the acquirement of gene promoter activity.

Results

Identification of 10 *pre-MIR058s* in *vitis*

In order to elucidate the origin and early evolutionary processes of *MIR058*, integrative sequence analysis of genome and transcriptome data from Vitaceae was carried out (Table S1). A total of 10 *pre-MIR058s* were identified from *V. vinifera*, *V. aestivalis*, *V. rotundifolia*, *V. pseudoreticulata*, *V. quinquangularis*, *V. amurensis*, and *V. riparia* x *V. rupestris* (Figure 2), all of which are members of the *Vitis* genus. No *pre-MIR058* was found in the genome and transcriptome data of other related genera in Vitaceae, such as *Ampelocissus*, the genus phylogenetically closest to *Vitis* [34]. It suggests that *pre-MIR058s* are specific to the *Vitis* genus. *Ampelocissus* and *Vitis* were split around 39.4 Ma in the late Eocene [34]. *Pre-MIR058s* could originate in a common ancestor of *Vitis* plants after that time. In addition, the *Vitis* genus includes two subgenera split around 37.3 Ma in the late Eocene [34]. *Pre-MIR058s* were identified from species of both subgenera, indicating the origination of *pre-MIR058s* before 37.3 Ma. Taken together, *pre-MIR058s* could originate around 37.3–39.4 Ma. It suggests that *pre-MIR058s* are evolutionarily



Figure 2. Hairpin structures of *MIR058* precursors from *Vitis*. It includes *V. vinifera* (Vvi), *V. aestivalis* (Vae), *V. rotundifolia* (Vro), *V. pseudoreticulata* (Vps), *V. quinqueangularis* (Vqu), *V. amurensis* var. *amurensis* from China (VaaC) and Russia (VaaR), *V. amurensis* var. *dissecta* (Vad), and *V. riparia* x *V. rupestris* (Vrr). Mature miRNA sequences are shown in red. Free energy of secondary structures calculated by the mfold program is shown as ΔG values in parentheses. Lower free energy indicates more stable secondary structure.

very young. It offers an opportunity to trace the early evolutionary processes of *pre-MIR058s* based on direct observation of the changes in DNA sequences.

MIR058 genes share high similarity to the whole antisense strand of *VviPPO1*

In order to elucidate the origin and early evolutionary process of *MIR058s*, I compared the sequences of *MIR058* genes and the target gene, *VviPPO1* (Figure 3A). The results showed a high similarity between *MIR058s* and the whole antisense strand of open reading frames (ORFs) of *VviPPO1*. The similarity not only exists in the *MIR058* precursor region but also in other parts of the gene, providing clear and convincing

evidence for the origination of *MIR058s* from an ancient *PPO*. Analysis of the genomic scaffolds and the whole genome sequence of *V. vinifera* identified four *VviPPOs*, all of which are single exon genes [35]. Phylogenetic analysis of *MIR058* genes and four *VviPPOs* showed that *MIR058s* had the highest similarity to the antisense strand of *VviPPO1* (Figure 3B). It indicates that *MIR058s* and *VviPPO1* share a common ancestor, or *MIR058a* were evolved from a copy of the ancient version of *VviPPO1*.

Origination of *pre-MIR058s* through the generation of short inverted repeats in a 18bp palindromic-like sequence

One of the significant characteristics of *MIRs* is that the *pre-MIR* contains an imperfect hairpin structure termed *MIRNA* precursor (*pre-MIR*) [1]. To determine how the hairpin-forming sequence of *MIR058s* originated, multiple sequence alignment of *pre-MIR058s* and *VviPPO1* was performed (Figure 3C). The results clearly showed that each of the *pre-MIR058s* has an inserted sequence with 31–36bp in length, which forms the 3'-arm of *MIR058* hairpin structure. The insertion is actually an inverted repeat (IR) of antisense *VviPPO1* sequence that corresponds to the 5'-arm of *MIR058*. It suggests that *pre-MIR058s* originated in a *PPO* through the generation of short IRs. Examination of the antisense strand of *VviPPO1* surrounding the IR generation site showed that *pre-MIR058s* were generated in a 18bp palindromic-like sequence flanked by 3bp direct repeats (DRs) (Figure 3C), which could be important for *MIR058* origination.

Deletion and mutation of *MIR058* gene sequences

MIRs are usually shorter than target genes and show low similarity to targets in regions other than *pre-MIRs* [2,12]. To investigate how it was caused, sequence comparison between *MIR058* genes and the antisense strand of *VviPPO1* ORF was conducted (Figure 3A). The results showed the existence of significant sequence variations in the 5'- and 3'-regions of *MIR058* genes (Figure 3A). The variations were caused mainly by short-deletions and mutations. Additionally, a large DNA fragment with approximately 868bp in length was lost in *V. aestivalis* *Vae-MIR058b* and *MIR058* genes from *V. pseudoreticulata*, *V. quinqueangularis*, *V. amurensis*, and *V. riparia* x *V. rupestris* (Figure 3A). It suggests that, during the early evolutionary process of plant *MIRs*, large- and short-deletions and mutations may occur. It could be the reason that caused *MIRs* to be shorter than target genes and showed low similarity to targets in regions other than *pre-MIRs*.

Identification and characterization of AasPPO-as-hp originating from an ancient *PPO* gene

Integrative sequence analysis of genome and transcriptome data from Vitaceae showed that *MIR058s* were specific to the *Vitis* genus. Is there a hairpin sequence originating from a *PPO* in other Vitaceae species? To address this question, I analysed all of the illumina and 454 RNA-seq data available for Vitaceae plants (Table S1) (<https://www.ncbi.nlm.nih.gov/sra>). The results showed that, although the Vitaceae species *Ampelocissus*

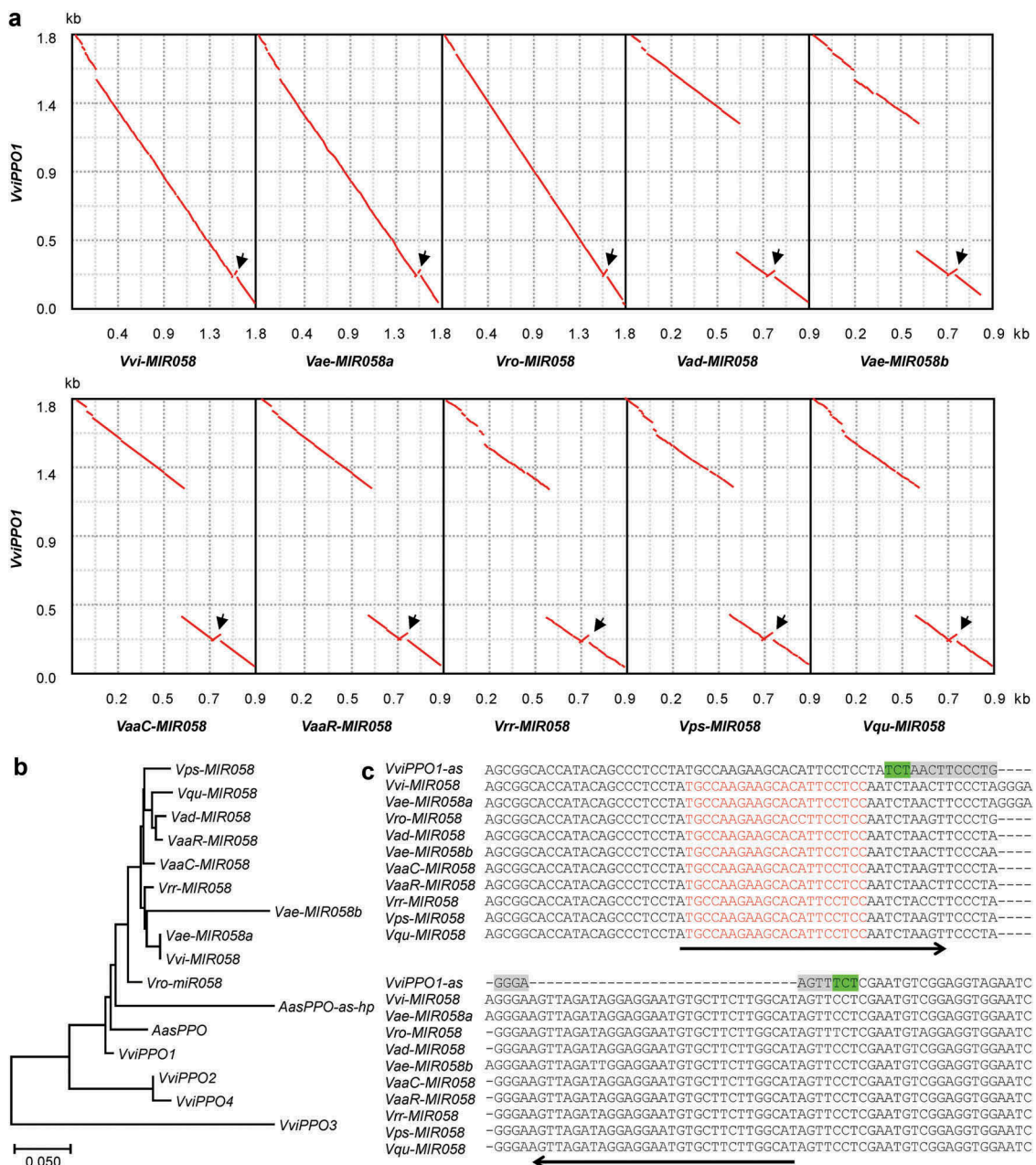


Figure 3. Formation of short IRs. (A) Dot-plots of *VviPPO1* ORF and *MIR058* sequences from *V. vinifera* (*Vvi*), *V. aestivalis* (*Vae*), *V. rotundifolia* (*Vro*), *V. amurensis* var. *dissecta* (*Vad*), *V. amurensis* var. *amurensis* from China (*VaaC*) and Russia (*VaaR*), *V. riparia* x *V. rupestris* (*Vrr*), *V. pseudoreticulata* (*Vps*), and *V. quinquangularis* (*Vqu*). The alignment and visualization were performed using zPicture (<http://zpicture.dcode.org/>). Arrows indicate the location of IRs formed. (B) Phylogenetic relationships of full-length *MIR058*s, *AasPPO-as-hp*, and the antisense strand of *VviPPO* ORFs and partial *VasPPO* ORF. (C) Alignment of *MIR058* sequences surrounding the IRs and the corresponding antisense sequence of *VviPPO1* (*VviPPO1-as*). Arrows indicate the IRs formed. The 18bp palindromic-like sequence (grey) flanked by 3bp DRs (green) is highlighted. Sequences for mature miR058 are shown in red.

ascendiflora did not contain *MIR058*, it had a genomic region showing high similarity to the antisense strand of *AasPPO* gene and contained a hairpin-forming sequence (Figures 4A and 4B). This region is termed *AasPPO-as-hp*. Phylogenetic analysis and sequence alignment of *AasPPO-as-hp* and *A. ascendiflora* *AAsPPO* revealed that *AasPPO-as-hp*, similar to the case of *MIR058*, originated from an ancient *PPO* gene (Figures 3B and 4B).

Sequence comparison showed that the generation of *AasPPO-as-hp* hairpin sequence is more complicated than *MIR058*. It includes deletion of a 1,143bp fragment and insertion

of a 55bp sequence at a site close to 5'-end of the open reading frame of *PPO* (Figures 4B and 4C). The inserted sequence contains 30bp IR of antisense *PPO* sequence (termed IR-1 hereafter) and 12bp DR of partial that *PPO* sequence (Figure 4B). Examination of the antisense strand of *AAsPPO* sequence identified a pair of 22–23bp imperfect IRs (termed IR-2 hereafter), one of which located at the 5'-junction of the deletion, whereas the other one located at the 3'-region of the hairpin-forming sequence (Figure 4B). The identified IR-2 could be important for the origination of the hairpin-forming sequence of *AasPPO-as-hp*. *A. ascendiflora* is an endangered plant species native to the

Russia and *V. riparia* x *V. rupestris*. Sequence comparison showed high similarity between the 5'-flanking region of *MIR058s* and the antisense strand of the 3'-flanking region of *VviPPO1* (Figure 5). It suggests that *MIR058* promoters are derived from the 3'-flanking region of a *PPO*. In addition to the conserved regions, I identified a region that existed in *MIR058* promoters but not in the *VviPPO1* promoter. This region is located at a site close to the corresponding stop codon of *VviPPO1* and contains multiple CuREs with the core sequence of GTAC (Figure 5). The CuRE region is featured by small size (e.g. 275bp in length for *Vvi-MIR058*)

and contains imperfect terminal inverted repeats (TIRs) and an internal AT-rich sequence (Figure 5). The insertion of this region in the 3'-flanking region of *PPO* generates 5bp target site duplications (TSDs) (Figure 5). Small size, TIR, internal AT-rich region and TSD are characteristics of MITE transposons [37]. It indicates that the gain of CuRE region is a result of MITE transposon insertion. The CuRE region was inserted in the upstream of the AT-rich TATA-box-like sequence. An inverted CAAT-box sequence was found at the 3'-end of the region (Figure 5). The capture of a multiple-CuRE-containing region in the early stage of *MIR058* origination indicates the

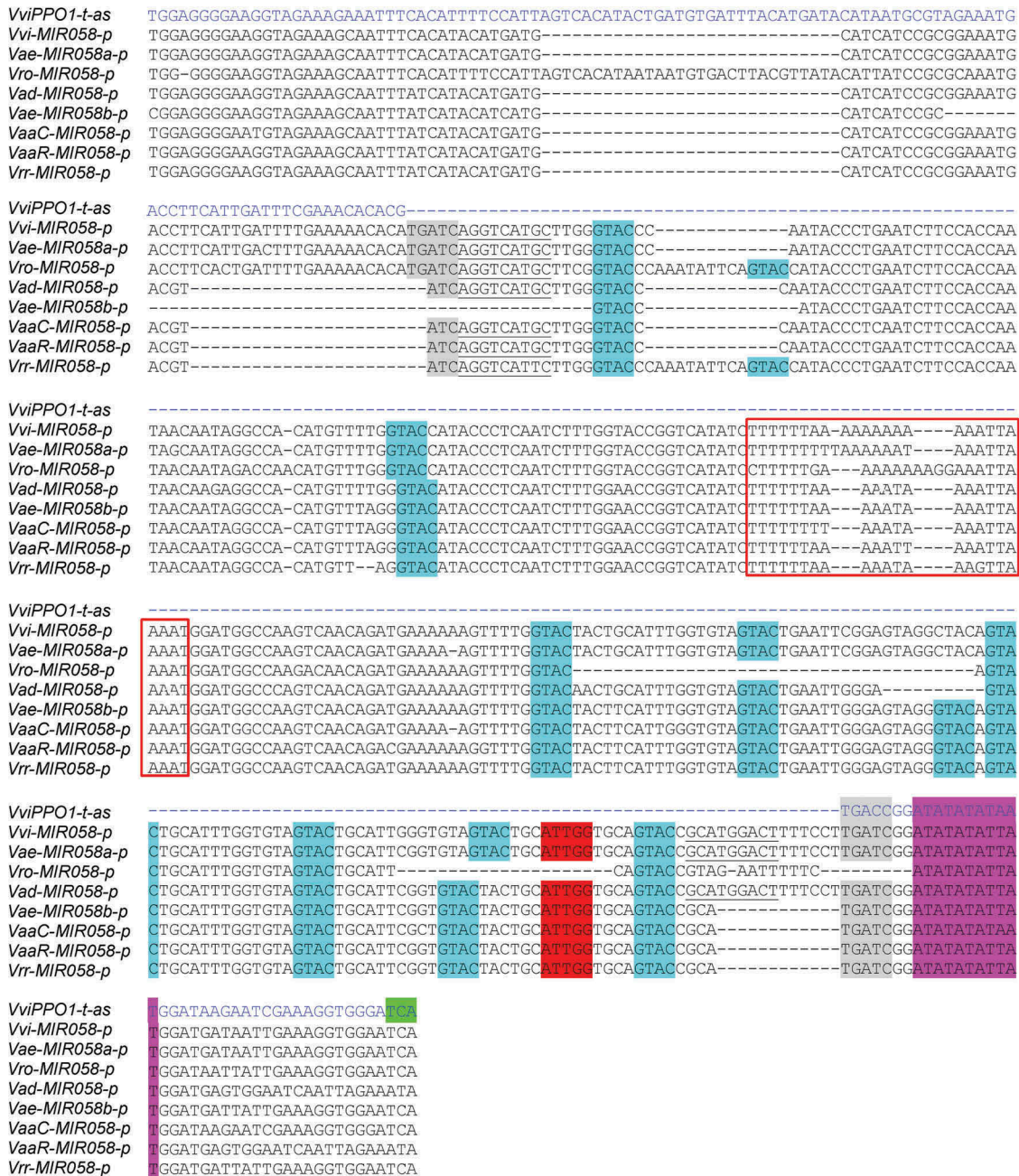


Figure 5. Gain of the CuRE regions. The 3'-flanking sequence of *VviPPO1* (antisense strand, *VviPPO1-t-as*) and the 5'-flanking sequences of *MIR058s* from *V. vinifera* (*Vvi*), *V. aestivalis* (*Vae*), *V. rotundifolia* (*Vro*), *V. amurensis* var. *dissecta* (*Vad*), *V. amurensis* var. *amurensis* from China (*VaaC*) and Russia (*VaaR*) and *V. riparia* x *V. rupestris* (*Vrr*) were aligned. The sequence of *VviPPO1-t-as* is shown in blue. CuREs are highlighted in pale blue. The antisense sequence of the stop codon of *VviPPO1* is highlighted in green. The 5bp TSDs are highlighted in grey. Imperfect TIRs of the inserted region are underlined. Internal TA-rich sequences of the inserted region are shown in a red box. The inverted CAAT-box sequence is highlighted in red. The AT-rich TATA-box-like region is highlighted in pink.

importance of Cu in the origination and evolution of Cu-responsive *MIRs*.

Promoter gain first, then pre-MIR058 and AasPPO-as-hp origination

I next asked the sequence of *pre-MIR* origination and promoter activity gain. To address this question, the 5'-flanking sequence of *AasPPO-as-hp* was identified and investigated. Sequence comparison showed that, similar to *MIR058*, the 5'-flanking sequence of *AasPPO-as-hp* contained a CuRE region (Figure 4D). This region includes four CuREs and shows high sequence similarity to parts of the CuRE region existing in *MIR058* promoters (Figures 4D and 5). Besides the CuRE region, high sequence similarity of *MIR058* and *AasPPO-as-hp* was also observed in the upstream of the CuRE region (Figure 4D). The existence of a common insertion sequence in the promoters of *MIR058* and *AasPPO-as-hp* indicates that this sequence was gained before the separation of *Vitis* and *Ampelocissus* plants. On the other hand, *pre-MIR058s* and the hairpin-forming sequences of *AasPPO-as-hp* are totally different, and the hairpin sequences of *AasPPO-as-hp* and *MIR058* were generated at different positions of the *PPO* gene (Figure 4C). The most rational explanation is that *pre-MIR058s* and *AasPPO-as-hp* originated after the separation of *Vitis* and *Ampelocissus*. Thus, the gain of the CuRE-containing sequence appears to occur before the origination of *pre-MIR058s* and the hairpin-forming sequences of *AasPPO-as-hp*. However, the possibility cannot be ruled out that *pre-MIR058s* and *AasPPO-as-hp* coexisted in an ancestor of *Vitis* and *Ampelocissus*, and after separation, *pre-MIR058s* was lost in *Ampelocissus* and *AasPPO-as-hp* was lost in *Vitis*. Even so, the origination of *pre-MIR058s* and *AasPPO-as-hp* in the ancestor could also happen after the gain of the CuRE-containing sequence.

In addition to the conserved region, sequence comparison of the 5'-flanking regions of *MIR058* and *AasPPO-as-hp* showed that a 930bp DNA fragment was lost in the 5'-flanking region of *AasPPO-as-hp* (Figure 4D). The loss of this fragment could happen in *Ampelocissus* after the separation of *Vitis* and *Ampelocissus*.

Origination and evolution of MIR1444 and MIR12112 genes

It has been shown that *MIR1444s* widely exist in *Populus*, *Salix* and *Idesia* [9–12]. *Populus* and *Idesia* have two *MIR1444* genes, termed *MIR1444a* and *MIR1444b*, respectively. Expansion of *MIR1444* genes in *Populus* and *Idesia* was through the Salicoid whole-genome duplication event [12,15]. *Salix* only contains *MIR1444b*. *MIR1444a* was lost in *Salix* through DNA segment deletion probably during chromosome rearrangements [12]. *Pre-MIR1444s* have extensive sequence similarity to *PPO* targets [12]. However, differing from *MIR058s* and *AasPPO-as-hp*, *MIR1444s* showed no conservation with *PPO* genes in DNA sequence other than the *pre-MIR1444* regions. It is consistent with the origination time of *pre-MIR1444s* and *pre-MIR058s*. *Pre-MIR1444s* originated in an ancient *PPO* gene before the Salicoid whole-genome duplication event happening 60 Ma [12,15], whereas *pre-MIR058* originated in an ancient *PPO*

gene around 37.3–39.4 Ma [34]. The results suggest that other parts of a *MIR* underwent less selective pressure than the *pre-MIR* part of the *MIR*. During origination and evolution of a *MIR*, the sequence other than *pre-MIR* may be altered significantly.

MIR12112 is another *PPO*-targeting *MIR*. It was only reported in *S. miltiorrhiza* [14]. In order to know the overall situation of *MIR12112* in plants, whole genome sequences and transcriptome data of *S. miltiorrhiza* and related plant species were systematically analysed. A total of 79 *pre-MIR12112s* were identified from species of the families Acanthaceae, Lentibulariaceae, Oleaceae, Orobanchaceae, Paulowniaceae, Pedaliaceae, Plantaginaceae, Verbenaceae, and Lamiaceae (Figure S1). All of these plant families belong to the order Lamiales. It suggests that *MIR12112s* are Lamiales-specific and relatively old *MIRs*. The number of *MIR12112* genes in a plant species varied significantly from 1 (e.g. species of the families Acanthaceae, Lentibulariaceae, Orobanchaceae and Paulowniaceae) to 6 (*Olea europaea* cv. *farga*), suggesting significant expansion of the *MIR12112* family in some plant species. The previous study had identified a *MIR12112* in *S. miltiorrhiza* [14]. In this study, two *Smi-MIR12112s* were identified. Thus, the previously reported *Smi-MIR12112* [14] is renamed *Smi-MIR12112a*, and the newly identified one is named *Smi-MIR12112b*.

Extensive gene number variation of *MIR12112* in Lamiales species could also result from genome duplication and DNA deletion events as the case of *MIR1444* [12]. A total of three *Sin-MIR12112* genes exist in *Sesamum indicum* (Figure S1). Examining the locations of *Sin-MIR12112s* in the genome assembly showed that *Sin-MIR12112s* were located at two homologous regions resulted from the sesame lineage-specific whole genome duplication event (Figure 6A) [38]. *Sin-MIR12112a* was located on LG6, whereas *Sin-MIR12112b* and *Sin-MIR12112c* were located on LG14. In addition to *Sin-MIR12112a*, a partial sequence of *Sin-MIR12112* was identified at a region close to *Sin-MIR12112a* (Figure 6A). The partial sequence contains partial mature miR12112 sequence and an intact miR12112* (data not shown). Similarly, there are three *OeuS-MIR12112s* in the *O. europaea* var. *sylvestris* genome assembly (Figure S1) [39]. *OeuS-MIR12112a* and *OeuS-MIR12112b* were located on pseudochromosome 6 at a region with homology to a pseudochromosome 21 region, where *OeuS-MIR12112c* was located [39]. From the draft genome of *O. europaea* cv. *farga* [40], a total of six *OeuF-MIR12112s* were identified (Figure S1). The reasons causing the increase of *OeuF-MIR12112* gene numbers remain to be elucidated. One possibility could be assemble errors resulted from genome heterozygosity.

Sequence comparison showed extensive sequence similarity between the stem portions of *Smi-MIR12112* precursors and partial of the antisense strand of *SmPPO* targets (Figure 6B). It indicated that *MIR12112* also originated from an ancient *PPO* gene as *MIR058*, *AasPPO-as-hp* and *MIR1444* did. The location of *MIR12112* precursors corresponds to 3'-end of the open reading frame of *PPO* (Figure 4C). Because of the relatively long evolutionary history of *MIR12112*, sequence information showing the birth of *MIR12112* genes from a *PPO* has been lost from the up- and down-streams of *pre-MIR12112s*. The loop regions of *MIR12112* precursors vary significantly with

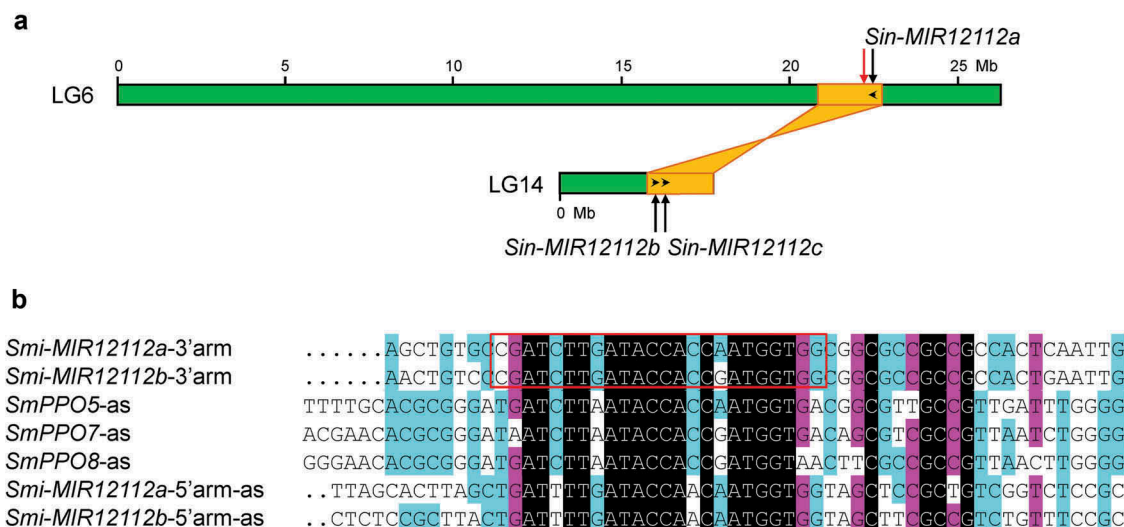


Figure 6. Origination and evolution of *MIR2112* genes. (A) Locations of *Sin-MIR2112s* (black arrows) and a partial sequence of *Sin-MIR2112* (red arrow) on pseudomolecules (LG6 and LG14, green boxes) of the *S. indicum* genome assembly. The schematic diagram of two homologous genome blocks (yellow) was adapted from Wang et al [47]. Transcription direction of *Sin-MIR2112s* is shown by black arrowheads. (B) Multiple sequence alignment of *S. miltiorrhiza* *Smi-MIR2112* precursors and *SmPPOs*. The sense orientation of *Smi-miR2112a* and *Smi-miR2112b* foldback arms, the antisense strand of *Smi-miR2112a** and *Smi-miR2112b** foldback arms, and parts of the antisense strand of *Smi-miR2112*-targeted *SmPPO5*, *SmPPO7* and *SmPPO8* are shown. The red box indicates mature *Smi-miR2112a* and *Smi-miR2112b* sequences.

frequent sequence insertion, deletion and mutation occurred (Figure S1). The results further suggested that the up- and down-streams, and the loop regions of *pre-MIR2112* underwent less evolutionary constraints than the stem regions.

Frequent birth and death of CuREs in promoters of *MIR2112*, *MIR1444* and *MIR058*

The existence of multiple CuREs in promoters of Cu-responsive *MIRs*, such as *A. thaliana* and *P. trichocarpa* *MIR397*, *MIR398* and *MIR408* and *P. trichocarpa* *MIR1444*, is important for their response to Cu availability [11,24–27]. In this study, I found that CuREs in *MIR058* and *AasPPOas-hp* promoters were gained by MITE transposon insertion (Figures 4D and 5). In order to obtain an overview of the origin and variation of CuREs in promoters of other *PPO*-targeting *MIRs*, I systematically investigated *MIR2112* and *MIR1444* gene promoters. A total of 34 *MIR2112* gene promoters from 19 plant species of the order Lamiales and six *MIR1444* gene promoters from *P. trichocarpa*, *P. Euphratica*, *Salix purpurea* and *S. Suchowensis* were identified (Figure 7). Sequence comparison showed the existence of CuREs in all of the *MIR2112* and *MIR1444* gene promoters investigated. The number of CuREs in *MIR2112* gene promoters varies significantly from 1 to 9 (Figure 7A). Changes of CuRE number were also observed in *MIR1444* promoters (Figure 7B). Taken together, the results from the investigation of *MIR2112* (Figure 7A), *MIR1444* (Figure 7B) and *MIR058* (Figure 4) gene promoters suggest frequent birth and death of CuREs. However, I am not able to trace the source of the CuRE region in promoters of *MIR2112* and *MIR1444*, since sequence information for the origin of CuREs in these *MIR* promoters has been lost. It is consistent with that *MIR2112* and *MIR1444* have a relatively long evolutionary history than *MIR058* and *AasPPOas-hp*. Even though the promoter

sequence has changed significantly in different plant species, the existence of CuREs in promoters of *PPO*-targeting *MIRs* suggests the importance of CuREs for *PPO*-targeting *MIRs*.

Discussion

A novel model for plant *MIR* origination and evolution

Although three models, including the inverted gene duplication [2,3], spontaneous evolution [6] and transposon transposition [7], have been proposed for the origination of plant *MIR* (Figure 1), there are many issues remained to be addressed, such as direct evidence for origin and early evolutionary process of hairpin sequences and promoters, the sequence of *MIR* origination and promoter activity gain, and the evolutionary force driven the origination and evolution of *MIRs*. Through systematic analysis of *PPO*-targeting *MIRs*, I propose, in this study, a novel model for *de novo* origination and evolution of plant *MIRs* (Figure 8). This model shows that at least some plant *MIRs* originated from an ancient version of targets through the generation of hairpin sequence in the ancient genes. I termed this model the short inverted repeat generation model. Although the exact mechanism of the generation of hairpin-forming sequence remains to be elucidated, a replication error occurring at pre-existing short inverted repeats in genomic sequences, known as Origin-Dependent Inverted-Repeat Amplification (ODIRA), was proposed for the formation of palindromic amplicons [41,42]. The hairpin-forming sequence in a *MIR* could be generated in an ancient target gene through a mechanism similar to ODIRA. Palindromic-like sequences (the 18-bp sequence in the case of *MIR058* precursor, Figure 3C) and imperfect IRs (the 22–23bp IR-2 in the case of *AasPPOas-hp* hairpin sequence, Figure 4B) in target genes might contribute to initiate the generation of hairpin sequences by causing errors during

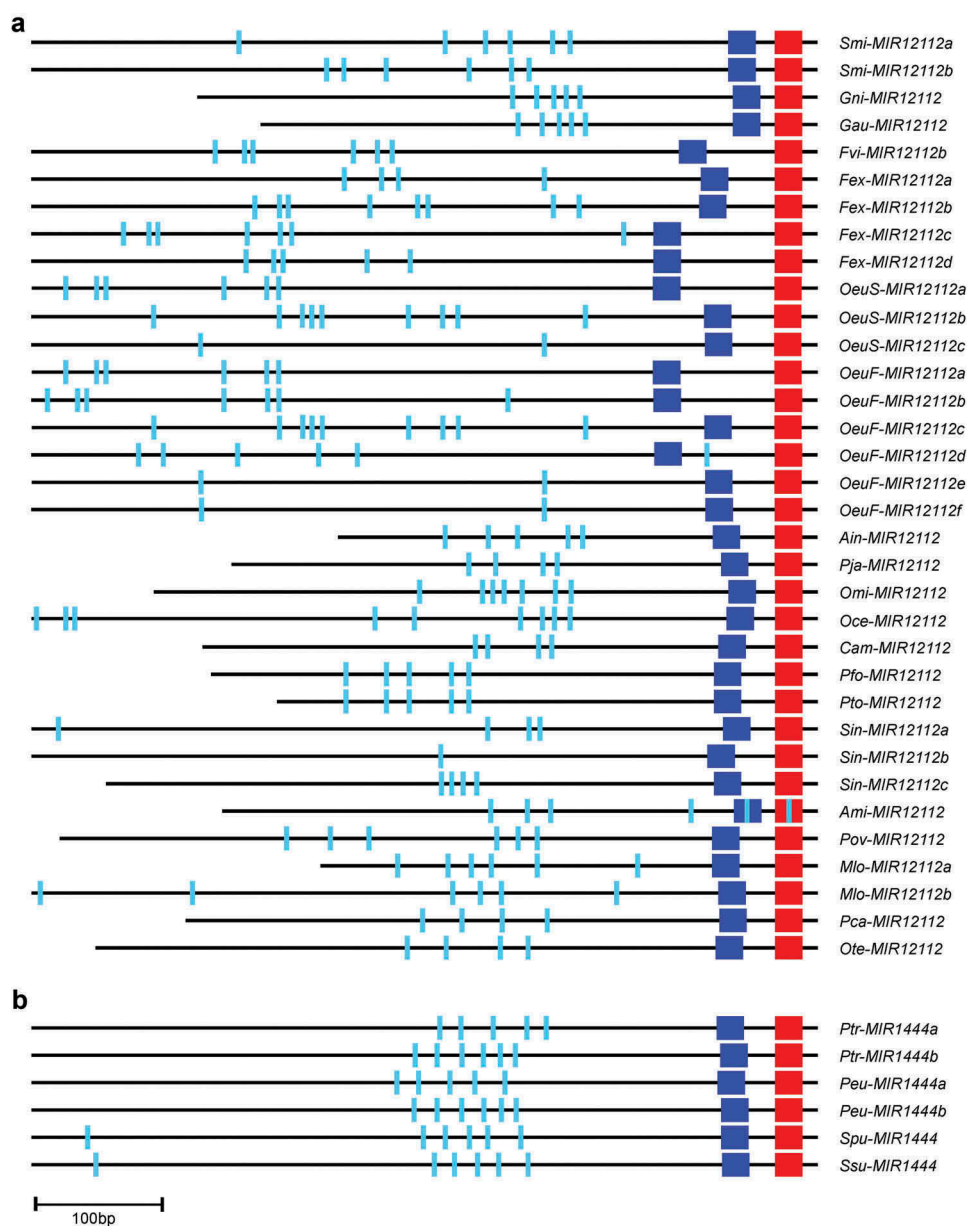


Figure 7. Distribution of CuREs in the 5'-flanking region of *MIR12112* and *MIR1444* genes. (A) Distribution of CuREs in the 5'-flanking region of *MIR12112* from *S. miltiorrhiza* (Smi), *G. nigrocaulis* (Gni), *G. aurea* (Gau), *Forsythia viridissima* (Fvi), *F. excelsior* (Fex), *O. europaea* cv. *sylvestris* (OeuS), *O. europaea* cv. *farga* (OeuF), *Aeginetia indica* (Ain), *Phtheirospermum japonicum* (Pja), *Orobancha minor* (Omi), *O. cernua* (Oce), *Conopholis americana* (Cam), *Paulownia fortunei* (Pfo), *P. tomentosa* (Pto), *S. indicum* (Sin), *Antirrhinum majus* (Ami), *Plantago ovata* (Pov), *M. longifolia* (Mlo), *Pogostemon cablin* (Pca), and *Ocimum tenuiflorum* (Ote). (B) Distribution of CuREs in the 5'-flanking region of *MIR1444* from *P. trichocarpa* (Ptr), *P. euphratica* (Peu), *S. purpurea* (Spu), and *S. Suchowensis* (Ssu). CuREs are indicated by pale blue vertical lines. The locations of mature miRNA sequence are shown by red boxes. The locations of miRNA* are indicated by blue boxes.

DNA duplication. In addition to the generation of hairpin sequence, I found that the promoter activity of *MIR058* and *AasPPO-as-hp* could be gained by insertion of a MITE sequence in the 3'-untranslated region of an ancient *PPO* gene. The MITE sequence contains significant *cis*-elements, such as CAAT-box and CuREs, and the insertion occurred in the upstream of the AT-rich TATA-box-like sequence before the generation of hairpin sequence (Figure 8). After the insertion of MITE and the generation of hairpin sequence, sequence deletion, insertion and mutation could occur during the evolution of a *MIR* (Figure 8).

Similarity and difference between the short inverted repeat generation model and the inverted gene duplication model

Both of the short inverted repeat generation model and the inverted gene duplication model show that *MIRs* were originated from target genes. The short inverted repeat generation model emphasizes that a *MIR* originated in an ancient target by generation of hairpin sequence or short inverted repeats (Figure 8), whereas the inverted gene duplication model suggests that a *MIR* originated by direct inverted duplication of a target gene, integration of a pseudogene-like sequence after reverse transcrip-

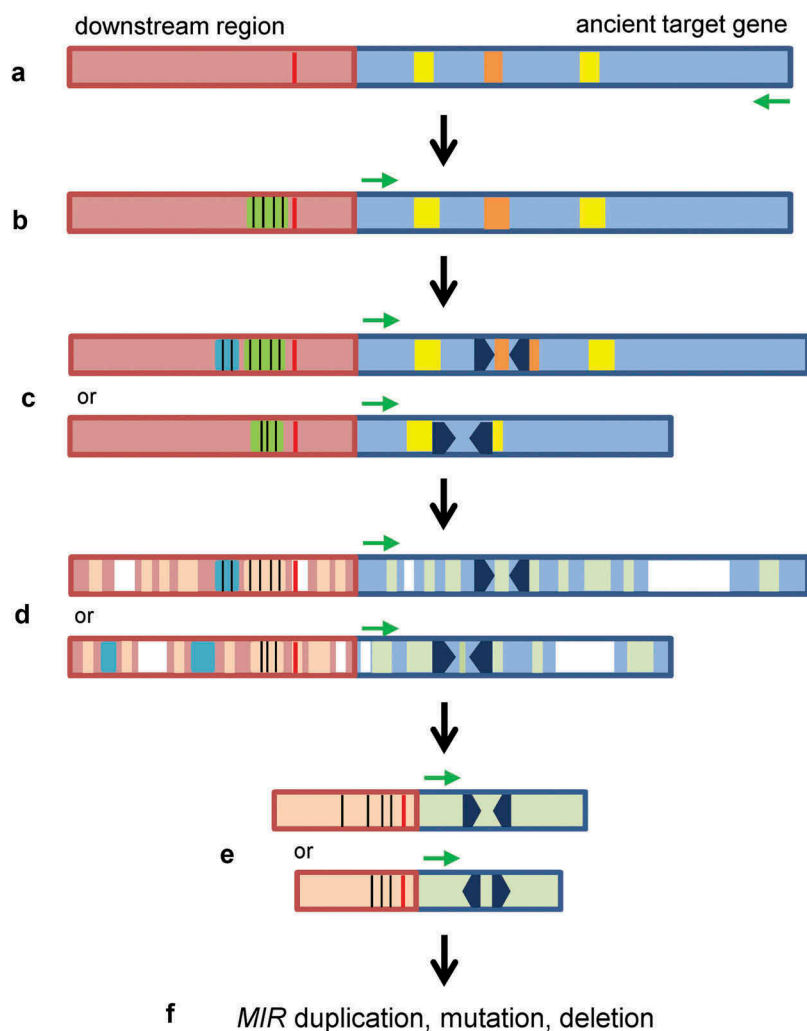


Figure 8. The short inverted repeat generation model for *de novo* origination and evolution of plant *MIRs*. (A) An ancient target gene (light blue) with palindromic-like (brown) and/or imperfect IR (yellow) sequences. Transcription direction is indicated by a green arrow. AT-rich TATA-box-like sequence is shown in red. (B) A MITE sequence (green) with significant *cis*-elements (black vertical lines), such as CAAT-box and CuREs, was inserted in the flanking region of the ancient target gene or a copy of the duplicated target gene. The insertion of MITE and subsequent evolution resulted in the gain of promoter activity. (C) Short inverted repeats (black arrowheads) were generated in the palindromic-like or imperfect IR sequences. *Cis*-elements could be dynamically changed (blue) in the promoters. (D) Deletion (white), insertion (blue) and mutation (light brown and light green) occurred during evolution. (E) Evolved *MIRs*. (F) The evolved *MIRs* could expand through duplication, diversify by mutation or lose by deletion during plant evolution.

tion, or juxtaposition of two closely related sequences from different members of a gene family (Figure 1) [1,2,43]. Generation of short inverted repeat in a target gene was proved by the results from *MIR058* and *AasPPO-as-hp*. Solid evidence to show the origination of *MIRs* through the three ways mentioned in the inverted gene duplication model is still lacking, and there is not completely satisfactory explanation for how the long hairpin sequence generated through inverted gene duplication become short hairpin sequence during *MIR* evolution. Through this study, I cannot rule out the possibility that some plant *MIRs* were originated by inverted gene duplication. In addition to the generation of hairpin sequence, the short inverted repeat generation model includes the mechanism of promoter activity gain and the sequence of *pre-MIR* generation and promoter activity gain, which are not mentioned in the inverted gene duplication model and other two models [2,6,7].

The short inverted repeat generation model was proposed based on the results from *PPO*-targeting *MIRs*, particularly *MIR058* and *AasPPO-as-hp*. It is very likely that this model can

be extended to some other evolutionarily young *MIRs* with extensive similarity or complementarity to target genes, and some old *MIRs* losing similarity to target genes after long-term evolution. However, due to the long evolutionary history of most plant *MIRs*, it is very difficult to precisely elucidate the origin and early evolutionary process of each *MIR* and to calculate the percentage of *MIRs* to be explained by each proposed model in a plant species. From this point of view, short inverted repeat generation, inverted gene duplication, spontaneous evolution and transposon transposition are four possible ways for plant *MIR* origination.

Significance of Cu in the origination and evolution of Cu-responsive MIRs

Cu is an essential micronutrient for plant growth and development. Maintenance of Cu at appropriate concentration in plant cells is important. Cu deficiency may cause abnormal growth, whereas Cu excess is toxic [11]. In order to adapt to

the changes of Cu availability in environments, plants have evolved a vital regulatory network in Cu uptake, distribution and molecular responses [11,44,45]. Cu-responsive miRNAs, such as miR397, miR398, miR408 and miR1444, are core components of this network. They are involved in the maintenance of Cu homeostasis in plant cells through regulating the expression of Cu-binding protein genes [11,25–27]. The promoters of these *MIRNA*s genes usually contain multiple CuREs, which sense Cu availability through SPL transcription factors [28–30]. In this study, I found that CuREs were captured by insertion of a CuRE-containing MITE sequence in the early stage of *MIR* origination and maintained by frequent birth and death during *MIR* evolution. The results indicate that Cu could be a significant force in driving the origination and evolution of *MIR*s targeting Cu-binding protein genes.

Materials and methods

Identification of *MIR058* and *AasPPO-as-hp* sequences

Vvi-MIR058 sequence was identified by blast analysis of *Vvi-MIR058* precursor against the whole genome sequence of *V. vinifera* using BLASTn [13,35,46]. *MIR058* sequences in other *Vitis* species were identified by analysis of *Vvi-MIR058* against next-generation sequence (NGS) data using BLASTn with word size set to be 7 [46]. Extension of partial *MIR058* sequences was performed through blast analysis against NGS data from the species. *MIR058* sequences identified from different species were cross-checked. CuRE region of *A. ascendiflora AasPPO-as-hp* was identified by analysis of the CuRE region of *Vvi-MIR058* promoter against *A. ascendiflora* NGS data using BLASTn with word size set to be 7 [46]. Full-length *A. ascendiflora AasPPO-as-hp* was obtained by extension of the CuRE sequence through blast analysis against the NGS data. *AasPPO* sequence was identified by analysis of *VviPPO1* against *A. ascendiflora* NGS data. Extension of partial *AasPPO* sequences was performed through blast analysis against NGS data from *A. ascendiflora*. The SRA accession numbers of illumina and 454 RNA-seq data used for *MIR058*, *AasPPO-as-hp* and *AasPPO* identification are listed in Table S1.

Identification of *MIR12112* sequences

Smi-MIR12112b sequence was identified by analysis of *S. miltiorrhiza Smi-MIR12112a* sequence against the whole genome sequence of *S. miltiorrhiza* using BLASTn [14,46,47]. *MIR12112* sequences in *Genlisea aurea*, *Fraxinus excelsior*, *O. europaea*, *S. indicum* and *Mentha longifolia* was identified by analysis of known *MIR12112* sequences against their whole genome sequence and the NGS data listed in Table S2 [38--40,46,48–50]. *MIR12112* sequences from plant species without the whole genome sequence available was identified and extended through analysis of known *MIR12112* sequences against NGS data using BLASTn with word size set to be 7 [46]. The SRA accession numbers of illumina and 454 RNA-seq data used for *MIR12112* identification are listed in Table S2.

Identification of the 5'-flanking sequence of *MIR1444s*

The 5'-flanking sequences of *MIR1444s* were identified through analysis of *MIR1444* precursors against the current genome assemblies of *P. trichocarpa* (v3.0), *P. euphratica* (v1.0), *S. purpurea* (v1.0, <https://phytozome.jgi.doe.gov/pz/portal.html>) and *S. suchowensis* (v1.0) using BLASTn [12,15,46,51,52].

Bioinformatics analysis and phylogenetic tree construction

Multiple sequence alignment was carried out using DNAMAN and M-Coffee [53]. The alignment and visualization of *MIR058s* and *VviPPO1* were performed using zPicture (<http://zpicture.dcode.org/>) [54]. The phylogenetic tree was constructed using MEGA7.0 by the neighbour-joining (NJ) method with 1000 bootstrap replicates [55].

RNA secondary structure prediction

Secondary structures of RNA sequence were predicted by the mfold program using the default parameters [56]. In each case, only the lowest energy structure was selected as described previously [33].

Accession numbers

Accession numbers for sequence data analysed are listed in Tables S1 and S2.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by the National Key Research and Development Program of China (grant number 2016YFD0600104), the CAMS Innovation Fund for Medical Sciences (CIFMS) (grant number 2016-I2M-3-016), and the National Natural Science Foundation of China (grant numbers 31570667, 81773836).

ORCID

Shanfa Lu  <http://orcid.org/0000-0001-6284-4389>

References

- [1] Voinnet O. Origin, biogenesis, and activity of plant microRNAs. *Cell*. 2009;136:669–687.
- [2] Allen E, Xie Z, Gustafson AM, et al. Evolution of microRNA genes by inverted duplication of target gene sequences in *Arabidopsis thaliana*. *Nat Genet*. 2004;36:1282–1290.
- [3] Fahlgren N, Howell MD, Kasschau KD, et al. High-throughput sequencing of *Arabidopsis* microRNAs: evidence for frequent birth and death of *MIRNA* genes. *PLoS One*. 2007;2:e219.
- [4] Xia R, Xu J, Arikiti S, et al. Extensive families of miRNAs and PHAS loci in Norway spruce demonstrate the origins of complex phasiRNA networks in seed plants. *Mol Biol Evol*. 2015;32:2905–2918.

- [5] Zhang Y, Xia R, Kuang H, et al. The diversification of plant *NBS-LRR* defense genes directs the evolution of microRNAs that target them. *Mol Biol Evol.* 2016;33:2692–2705.
- [6] de Felippes FF, Schneeberger K, Dezulian T, et al. Evolution of *Arabidopsis thaliana* microRNAs from random sequences. *RNA.* 2008;14:2455–2459.
- [7] Piriyaopongsa J, Jordan IK. Dual coding of siRNAs and miRNAs by plant transposable elements. *RNA.* 2008;14:814–821.
- [8] Xia R, Meyers BC, Liu Z, et al. MicroRNA superfamilies descended from miR390 and their roles in secondary small interfering RNA biogenesis in eudicots. *Plant Cell.* 2013;25:1555–1572.
- [9] Lu S, Sun YH, Chiang VL. Stress-responsive microRNAs in *Populus*. *Plant J.* 2008;55:131–151.
- [10] Lu S, Sun YH, Chiang VL. Adenylation of plant miRNAs. *Nucleic Acids Res.* 2009;37:1878–1885.
- [11] Lu S, Yang C, Chiang VL. Conservation and diversity of microRNA associated copper-regulatory networks in *Populus trichocarpa*. *J Integr Plant Biol.* 2011;53:879–891.
- [12] Wang M, Li C, Lu S. Origin and evolution of *MIR1444* genes in Salicaceae. *Sci Rep.* 2017;7:39740.
- [13] Ren G, Wang B, Zhu X, et al. Cloning, expression, and characterization of miR058 and its target *PPO*, during the development of grapevine berry stone. *Gene.* 2014;548:166–173.
- [14] Li C, Li D, Li J, et al. Characterization of the polyphenol oxidase gene family reveals a novel microRNA involved in posttranscriptional regulation of *PPOs* in *Salvia miltiorrhiza*. *Sci Rep.* 2017;7:44622.
- [15] Tuskan GA, Difazio S, Jansson S, et al. The genome of black cottonwood *Populus trichocarpa* (Torr. & Gray). *Science.* 2006;313:1596–1604.
- [16] Wang C, Leng X, Zhang Y, et al. Transcriptome-wide analysis of dynamic variations in regulation modes of grapevine microRNAs on their target genes during grapevine development. *Plant Mol Biol.* 2014;84:269–285.
- [17] Kozomara A, Griffiths-Jones S. miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res.* 2014;42(Database issue):D68–73.
- [18] Ma Y, Yuan L, Wu B, et al. Genome-wide identification and characterization of novel genes involved in terpenoid biosynthesis in *Salvia miltiorrhiza*. *J Exp Bot.* 2012;63:2809–2823.
- [19] Liu M, Lu S. Plastoquinone and ubiquinone in plants: biosynthesis, physiological function and metabolic engineering. *Front Plant Sci.* 2016;7:1898.
- [20] Deng Y, Lu S. Biosynthesis and regulation of phenylpropanoids in plants. *Crit Rev Plant Sci.* 2017;36:257–290.
- [21] Zhang L, Lu S. Overview of medicinally important diterpenoids derived from plastids. *Mini-Rev Med Chem.* 2017;17:988–1001.
- [22] Quinn JM, Merchant S. Two copper-responsive elements associated with the *Chlamydomonas* *Cyc6* gene function as targets for transcriptional activators. *Plant Cell.* 1995;7:623–628.
- [23] Quinn JM, Barraco P, Eriksson M, et al. Coordinate copper- and oxygen-responsive *Cyc6* and *Cpx1* expression in *Chlamydomonas* is mediated by the same element. *J Biol Chem.* 2000;275:6080–6089.
- [24] Nagae M, Nakata M, Takahashi Y. Identification of negative *cis*-acting elements in response to copper in the chloroplastic iron superoxide dismutase gene of the moss *Barbula unguiculata*. *Plant Physiol.* 2008;146:1687–1696.
- [25] Sunkar R, Kapoor A, Zhu JK. Posttranscriptional induction of two Cu/Zn superoxide dismutase genes in *Arabidopsis* is mediated by downregulation of miR398 and important for oxidative stress tolerance. *Plant Cell.* 2006;18:2051–2065.
- [26] Yamasaki H, Abdel-Ghany SE, Cohu CM, et al. Regulation of copper homeostasis by micro-RNA in *Arabidopsis*. *J Biol Chem.* 2007;282:16369–16378.
- [27] Abdel-Ghany SE, Pilon M. MicroRNA-mediated systemic downregulation of copper protein expression in response to low copper availability in *Arabidopsis*. *J Biol Chem.* 2008;283:15932–15945.
- [28] Yamasaki H, Hayashi M, Fukazawa M, et al. SQUAMOSA promoter binding protein-like7 is a central regulator for copper homeostasis in *Arabidopsis*. *Plant Cell.* 2009;21:347–361.
- [29] Kropat J, Tottey S, Birkenbihl RP, et al. A regulator of nutritional copper signaling in *Chlamydomonas* is an SBP domain protein that recognizes the GTAC core of copper response element. *Proc Natl Acad Sci USA.* 2005;102:18730–18735.
- [30] Sommer F, Kropat J, Malasarn D, et al. The CRR1 nutritional copper sensor in *Chlamydomonas* contains two distinct metal-responsive domains. *Plant Cell.* 2010;22:4098–4113.
- [31] Lu S, Li Q, Wei H, et al. Ptr-miR397a is a negative regulator of laccase genes affecting lignin content in *Populus trichocarpa*. *Proc Natl Acad Sci USA.* 2013;110:10848–10853.
- [32] Lu S, Sun YH, Amerson H, et al. MicroRNAs in loblolly pine (*Pinus taeda* L.) and their association with fusiform rust gall development. *Plant J.* 2007;51:1077–1098.
- [33] Lu S, Sun YH, Shi R, et al. Novel and mechanical stress-responsive microRNAs in *Populus trichocarpa* that are absent from *Arabidopsis*. *Plant Cell.* 2005;17:2186–2203.
- [34] Liu XQ, Ickert-Bond SM, Nie ZL, et al. Phylogeny of the *Ampelocissus-Vitis* clade in vitaceae supports the new world origin of the grape genus. *Mol Phylogenet Evol.* 2016;95:217–228.
- [35] Jaillon O, Aury JM, Noel B, et al. The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature.* 2007;449:463–467.
- [36] Yeo CK, Ang WF, Lok AFSL, et al. The conservation status of *Ampelocissus* planch. (Vitaceae) of Singapore, with a special note on *Ampelocissus ascendiflora* latiff. *Nat Singapore.* 2013;6:45–53.
- [37] Feschotte C, Jiang N, Wessler SR. Plant transposable elements: where genetics meets genomics. *Nat Rev Genet.* 2002;3:329–341.
- [38] Wang L, Yu S, Tong C, et al. Genome sequencing of the high oil crop sesame provides insight into oil biosynthesis. *Genome Biol.* 2014;15:R39.
- [39] Unver T, Wu Z, Sterck L, et al. Genome of wild olive and the evolution of oil biosynthesis. *Proc Natl Acad Sci U S A.* 2017;114: E9413–9422.
- [40] Cruz F, Julca I, Gómez-Garrido J, et al. Genome sequence of the olive tree, *Olea europaea*. *Gigascience.* 2016;5:29.
- [41] Brewer BJ, Payen C, Raghuraman MK, et al. Origin-dependent inverted-repeat amplification: a replication-based model for generating palindromic amplicons. *PLoS Genet.* 2011;7:e1002016.
- [42] Brewer BJ, Payen C, Di Rienzi SC, et al. Origin-dependent inverted-repeat amplification: tests of a model for inverted DNA amplification. *PLoS Genet.* 2015;11:e1005699.
- [43] Cui J, You C, Chen X. The evolution of microRNAs in plants. *Curr Opin Plant Biol.* 2016;35:61–67.
- [44] Grotz N, Guerinet ML. Molecular aspects of Cu, Fe and Zn homeostasis in plants. *Biochim Biophys Acta.* 2006;1763:595–608.
- [45] Puig S, Andres-Colas N, Garcia-Molina A, et al. Copper and iron homeostasis in *Arabidopsis*: responses to metal deficiencies, interactions and biotechnological applications. *Plant Cell Environ.* 2007;30:271–290.
- [46] Altschul SF, Madden TL, Schäffer AA, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 1997;25:3389–3402.
- [47] Xu H, Song J, Luo H, et al. Analysis of the genome sequence of the medicinal plant *Salvia miltiorrhiza*. *Mol Plant.* 2016;9:949–952.
- [48] Leushkin EV, Sutormin RA, Nabieva ER, et al. The miniature genome of a carnivorous plant *Genlisea aurea* contains a low number of genes and short non-coding sequences. *BMC Genomics.* 2013;14:476.
- [49] Sollars ES, Harper AL, Kelly LJ, et al. Genome sequence and genetic diversity of European ash trees. *Nature.* 2017;541:212–216.
- [50] Vining KJ, Johnson SR, Ahkami A, et al. Draft genome sequence of *Mentha longifolia* and development of resources for mint cultivar improvement. *Mol Plant.* 2017;10:323–339.
- [51] Ma T, Wang J, Zhou G, et al. Genomic insights into salt adaptation in a desert poplar. *Nat Commun.* 2013;4:2797.

- [52] Dai X, Hu Q, Cai Q, et al. The willow genome and divergent evolution from poplar after the common genome duplication. *Cell Res.* 2014;24:1274–1277.
- [53] Di Tommaso P, Moretti S, Xenarios I, et al. T-Coffee: a web server for the multiple sequence alignment of protein and RNA sequences using structural information and homology extension. *Nucleic Acids Res.* 2011;39(Web Server issue): W13–17.
- [54] Ovcharenko I, Loots GG, Hardison RC, et al. zPicture: dynamic alignment and visualization tool for analyzing conservation profiles. *Genome Res.* 2004;14:472–477.
- [55] Kumar S, Stecher G, Tamura K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol Biol Evol.* 2016;33:1870–1874.
- [56] Zuker M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* 2003;31:3406–3415.