Miniature Inverted-repeat Transposable Elements Drive Rapid MicroRNA Diversification in Angiosperms

Zhonglong Guo, Zheng Kuang, Yihan Tao, Haotian Wang, Miaomiao Wan, Chen Hao, Fei Shen, Xiaozeng Yang, and Lei Li

State Key Laboratory of Protein and Plant Gene Research, School of Advanced Agricultural Sciences and School of Life Sciences, Peking University, Beijing 100871, China
Co-Innovation Center for Sustainable Forestry in Southern China, College of Biology and the Environment, Nanjing Forestry University, Nanjing 210037, China
Peking-Tsinghua Center for Life Sciences, Academy for Advanced Interdisciplinary Studies, Peking University, Beijing 100871, China
State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming 650223, China
Beijing Agro-biotechnology Research Center, Beijing Academy of Agriculture and Forestry Sciences, Beijing 100097, China

*Corresponding authors: E-mails: lei.li@pku.edu.cn; yangxiaozeng@baafs.net.cn.
Associate editor: Irina Arkhipova

Abstract

MicroRNAs (miRNAs) are fast evolving endogenous small RNAs that regulate organism function and behavior in both animals and plants. Although models for de novo miRNA biogenesis have been proposed, the genomic mechanisms driving swift diversification of the miRNA repertoires in plants remain elusive. Here, by comprehensively analyzing 21 phylogenetically representative plant species, ranging from green algae to angiosperms, we systematically identified de novo miRNA events associated with 8,649 miRNA loci. We found that 399 (4.6%), 466 (5.4%), and 1,402 (16.2%) miRNAs were derived from inverted gene duplication events, long terminal repeats of retrotransposons, and miniature inverted-repeat transposable elements (MITEs), respectively. Among the miRNAs of these origins, MITEs, especially those belonging to the Mutator, Tc1/Mariner, and PIF/Harbinger superfamilies, were the predominant genomic source for de novo miRNAs in the 15 examined angiosperms but not in the six non-angiosperms. Our data further illustrated a transposition–transcription process by which MITEs are converted into new miRNAs (termed MITE-miRNAs) whereby properly sized MITEs are transcribed and therefore become potential substrates for the miRNA processing machinery by transposing into introns of active genes. By analyzing the 58,038 putative target genes for the 8,095 miRNAs, we found that the target genes of MITE-miRNAs were preferentially associated with response to environmental stimuli such as temperature, suggesting that MITE-miRNAs are pertinent to plant adaptation. Collectively, these findings demonstrate that molecular conversion of MITEs is a genomic mechanism leading to rapid and continuous changes to the miRNA repertoires in angiosperms.

Key words: plant microRNAs, MITEs, origination, evolution.

Introduction

MicroRNAs (miRNAs) are an important class of endogenous regulatory RNAs in eukaryotes regulating gene expression. The 20–24 nucleotide miRNAs arise from precursor transcripts called pre-miRNAs that contain sequences capable of forming the characteristic intra-molecular hairpin structures (Voinnet 2009; Yang et al. 2012; Rogers and Chen 2013; Bartel 2018). Processed by the evolutionarily conserved double-stranded RNA-specific endoribonuclease (Dicer or Dicer-like in plants), the yielded mature miRNAs guide both transcriptional and post-transcriptional gene regulations (Bartel 2004; Voinnet 2009; Rogers and Chen 2013). In plants, numerous studies have established the regulatory function of miRNAs in governing development and stress responses (Voinnet 2009; Rogers and Chen 2013; Rodriguez et al. 2016; Gao et al. 2021). However, the vast plant miRNA repertoires exhibit a high level of genetic diversity with only a fraction conserved at different taxonomic levels (Cuperus et al. 2011; Chavez Montes et al. 2014; Guo et al. 2020). Thus, genomic mechanisms for swift de novo miRNA biogenesis must be in place, which, when coupled with selection and gene loss, would allow individual lineages to retain largely non-overlapping sets of miRNAs (Chen and Rajewsky 2007; Moran et al. 2017; Zhang et al. 2021).

In plants, three models of de novo miRNA biogenesis have been conceptualized. The target gene inverted duplication model proposes that occasional inverted duplication during gene family expansion might yield transcripts...
that fold into near-perfect hairpin structures, which are initially processed to produce small interfering RNAs (siRNAs). When mutations compromise complementarity of the hairpin structures, they could switch to produce miRNAs (referred to as TID-miRNAs herein) (Allen et al. 2004; Cui et al. 2017; Baldrich et al. 2018). Characterization of miRNAs in representative species of the phylum Cnidaria showed that miRNA precursors can originate from their own targets, suggesting that TID-miRNAs represent an ancestral mechanism utilized by both plants and early animals (Fridrich et al. 2020). The long terminal repeat (LTR) model predicts that consecutive retrotransposons in opposite configurations may produce transcripts derived from readthrough events. These LTR-containing transcripts could fold into long hairpin structures triggering the production of siRNAs and some might eventually give rise to miRNAs (referred to as LTR-miRNAs; Piriyapongsa and Jordan 2008; Voinnet 2009). The third model is based on miniature inverted-repeat transposable elements (MITEs), which are truncated derivatives of autonomous DNA transposons and contain terminal inverted repeats (TIRs) flanked by two short direct repeats called target site duplications (TSDs; Feschotte and Mouchès 2000; Yang et al. 2001; Feschotte et al. 2003). Their relatively small sizes (generally 50–800 bp long) and high copy number have led to the proposal that MITEs co-transcribed with the host genes could produce the imperfect hairpin structures for biogenesis of new miRNAs (termed MITE-miRNAs; Piriyapongsa and Jordan 2007; Kuang et al. 2009; Lu et al. 2012; Cui et al. 2017). More recently, involvement of MITE-derived miRNAs in wheat immune response to pathogens was observed (Poretti et al. 2019). However, these miRNA biogenesis models have not been assessed on a genome scale in different plant species.

In this study, by comprehensively integrating information of the miRNA portfolios with genomic annotations, we systemically assessed the three miRNA biogenesis models in 21 phylogenetically representative plant species. The overarching findings illustrated a transposition–transcription process for MITE-miRNA biogenesis and demonstrated MITEs as the predominant genomic source for de novo miRNAs with recognizable origins in angiosperm. We further found that the MITE-miRNAs were selected for targeting genes preferentially associated with environmental plasticity. These findings add new insights into plant miRNA evolution.

Results

Plant MiRNAs are Highly Diversified

Both miRNAs and siRNAs are functional small RNAs that share some biogenesis features (Moran et al. 2017). To focus on miRNAs, we employed a specifically annotated data set of 20,338 miRNAs in 88 plant species that are phylogenetically ranging from chlorophytes to angiosperms (Guo et al. 2020; supplementary table S1, Supplementary Material online). Analysis of the miRNA data set identified 5,022 miRNA families based on homology of the mature miRNAs (supplementary fig. S1A, Supplementary Material online). We found that none of the 249 identified algal miRNA families were present in land plants (supplementary fig. S1B and table S2, Supplementary Material online). Pairwise comparison of miRNA families among 30 taxonomic families of land plants showed that at least 40% miRNA families were present only in a single family (supplementary fig. S1C, Supplementary Material online). In four genera of the Poaceae family, we identified 1,073 miRNA families of which only 27 (2.5%) were present in all four genera, whereas 1,016 (94.7%) were present in a specific genus (supplementary fig. S1D and table S3, Supplementary Material online). Within the Oryza genus, 202 (41.4%) of the 488 miRNA families in modern Oryza sativa ssp. japonica were not present in the other seven examined species or subspecies (fig. 1A).

We traced the rate of net gain of de novo miRNA families in the Oryza genus over evolutionary time (supplementary fig. S2, Supplementary Material online). By examining synteny between closely related Oryza genomes and reciprocal-best whole-genome alignments, 175 de novo protein-coding genes were identified in O. sativa
spp. japonica after its divergence from Oryza punctata ~3.4 million years (My) ago (Zhang et al. 2019). By the same comparison, we identified 365 de novo miRNA families in O. sativa spp. japonica, a net gain of 107.4 new miRNA families per My (fig. 1B), or 2.1 times higher than the net gain of protein-coding genes from ancestral non-coding sequences (Zhang et al. 2019). After the most recent divergence from Oryza rufipogon, the estimated rate of net gain was 255.4 miRNA families per My or 5.5 times higher than de novo protein-coding genes (fig. 1C). These observations support previous conclusions of high divergence rates of plant miRNAs (Cuperus et al. 2011; Chavez Montes et al. 2014; Guo et al. 2020), which we further attributed to a rapid origination of novel miRNA families.

MITE is the Dominant Source for de novo miRNAs in Angiosperm

To identify the genomic sources for de novo miRNAs, we selected 21 phylogenetically representative species with high-quality reference genomes and available sRNA-Seq data, including two green algae (Chlamydomonas reinhardtii and Volvox carteri), one moss (Physcomitrella patens), one lycophyte (Selaginella moellendorfii), two gymnosperms (Ginkgo biloba and Picea abies), one basal angiosperm (Amborella trichopoda), five monocots in Poaceae, and nine dicotyledons from five taxonomic families (supplementary table S4, Supplementary Material online). In these species, we compared the miRNAs against siRNAs in the sRNAanno database (Chen et al. 2021) and found that 89.2% miRNA loci do not overlap with the siRNAs (supplementary table S5, Supplementary Material online). Three species, Brachypodium distachyon, O. sativa, and Solanum tuberosum, were noticed to have significant overlapping between the annotated miRNAs and siRNAs (supplementary table S5, Supplementary Material online). This could be due to uneven annotation standards for the small RNAs across different species or variations in the number of loci exhibiting features that qualify them as both a miRNA and a siRNA by current methods (supplementary table S6, Supplementary Material online).

Fitting the 8,649 miRNA loci in the 21 species against the three de novo biogenesis models using stringent criteria (supplementary fig. S3, Supplementary Material online), we identified 399 TID-miRNAs (4.6%) belonging to 293 families, 466 LTR-miRNAs (5.4%) belonging to 193 families, and 1,402 MITE-miRNAs (16.2%) belonging to 556 families (supplementary fig. S4, table S7, and data S1–S5, Supplementary Material online). The three types of miRNAs were found in both green algae and land plant lineages (fig. 2A and B, supplementary fig. S5, Supplementary Material online). Further, most TID-miRNA families (83.6%), LTR-miRNA families (81.9%), and MITE-miRNA families (88.3%) were species specific, indicating that novel miRNAs have heterogeneous origins in plants. Among the three types of miRNAs, MITE-miRNAs exhibited the highest numbers, due to a drastic expansion in angiosperms (fig. 2A and B). Only six, or <0.5% miRNAs were MITE-miRNAs in the six examined non-angiosperm plants: three in V. carteri, two in C. reinhardtii, one in Ph. patens, and none in the gymnosperm G. biloba and Pi. abies (fig. 2B, supplementary fig. S5, Supplementary Material online). In contrast, 1,396 (20.6%) miRNAs, or 551 (16.9%) miRNA families, were identified as MITE-miRNAs in the 15 examined angiosperms (supplementary fig. S5, Supplementary Material online), or a burst of two orders of magnitude over non-angiosperms. In the basally branching angiosperm A. trichopoda, 59 or 25.0% miRNAs (34 or 23.3% families) were MITE-miRNAs, suggesting that the burst of MITE-miRNAs had already occurred in early angiosperms (fig. 2B, supplementary fig. S5, Supplementary Material online).

We also increased the annotation standards for MITEs by including homology to known DNA transposons (Feschotte et al. 2003; Bao et al. 2015), which identified 600 MITE-miRNAs with higher confidence (supplementary table S7 and data S2, Supplementary Material online). In all subsequent analyses, we found that conclusions drawn from using the 1,402 MITE-miRNAs were consistent with those using the 600 MITE-miRNAs (supplementary data S6, Supplementary Material online). To be more comprehensive, we report the findings based on the 1,402 MITE-miRNAs. The miR437 family, which is conserved in Poaceae (Sunkar et al. 2005; Sunkar and Jagadeeswaran 2008), was used as an illustrative example for the MITE-miRNAs. In the five analyzed species, 32 of the 36 MIR437 loci, including 20 of the 23 loci in Sorghum bicolor, were identified as MITE-miRNAs (supplementary fig. S6A). The corresponding MITEs were homologous to the DNA transposon STOWAWAY14_SB (supplementary data S2, Supplementary Material online). Phylogeny and homology analysis further showed that 11 loci in Sor. bicolor have formed by transposition from independent MITEs before the divergence from Zea mays, whereas nine were specifically formed in Sor. bicolor (supplementary fig. S6B, Supplementary Material online). Sbi-MIR437X and Sbi-MIR437M were used as representative examples to showcase the conservation of the MITE features. Sequence alignment showed that both miR437 and miR437 star in Sbi-MIR437X and Sbi-MIR437M locate in regions homologous to the TIR of the MITEs, whereas the 2 bp "TA" TSDs are still present at both ends of the miRNA loci (supplementary fig. S7, Supplementary Material online).

We compared the rates of net gain of the three types of miRNAs in the Oryza genus. In O. sativa, we identified 44 TID-miRNA, 10 LTR-miRNA, and 118 MITE-miRNA families (fig. 2B, supplementary fig. S8A, Supplementary Material online). We found that the rates of net gain of MITE-miRNA families were consistently higher than LTR-miRNA and TID-miRNA families throughout evolution of the Oryza genus. Overall, our findings demonstrate that MITEs are the dominant source for rapid de novo miRNA biogenesis in Oryza, a conclusion corroborated by analyzing miRNA origination events in the Solanum genus (supplementary fig. S8B, Supplementary Material online).
Diversification of MITE Families Influences the Frequency of MITE-miRNAs

Within the examined angiosperm species, no correlation was found between the number of MITE-miRNA families and the number of MITEs (supplementary table S8, Supplementary Material online; Pearson’s $r = 0.005$, $P = 0.767$). This observation prompted us to examine the MITE families. Based on homology of the TIR and TSD, MITEs in the green plant lineage are divided into seven superfamilies, including Tc1/Mariner, PIF/Harbinger, Mutator, CACTA, hAT, P element, and Novosib (Feschotte and Mouchès 2000; Fattash et al. 2013; Chen et al. 2014). The examined plants exhibited drastically different composition of the seven superfamilies (supplementary fig. S9 and table S9, Supplementary Material online).

Consistent with previous reports (Chen et al. 2014), P element and Novosib were only found in C. reinhardtii and thus had little impact on the MITE-miRNAs. In the other five superfamilies, the proportion of MITEs and the proportion of MITE-miRNA families derived from each superfamily showed a strong correlation (fig. 2C). Diameter of the circles represents Pearson’s correlation coefficient with the size of the box set to one. *$P < 0.05$, **$P < 0.01$. (D–F) Proportions of MITE-miRNA families derived from the five MITE superfamilies in all examined angiosperm species (D), examined dicots (E), and individual species (F). (G) Average proportions of MITE-miRNA families derived from the five MITE superfamilies in the six taxonomic groups as shaded.
angiosperm was derived from Mutator (43.4%), Tc1/Mariner (22.7%), and PIF/Harbinger (16.2% [fig. 2D, supplementary tables S11 and S12, Supplementary Material online]). Furthermore, over half MITE-miRNA families (59.0%) in dicots were derived from Mutator alone (fig. 2E). In the examined monocots, 39.4% MITE-miRNA families in the clade consisting of O. sativa, B. distachyon, and Setaria italica were derived from Tc1/Mariner whilst 50% MITE-miRNA families in the clade containing Sor. bicolor and Z. mays were from PIF/Harbinger (fig. 2F, G).

Within each superfamily, MITEs of sequence homology were further grouped into families. After annotating the MITE families, we found that the number of MITE families in angiosperm was generally an order of magnitude higher than that in non-angiosperm. Comparing the number of MITE families and MITE-miRNA families revealed a significant positive linear correlation (Pearson’s \( r = 0.546, P < 0.01; \) supplementary fig. S10, Supplementary Material online). We found that 70.7% MITE families in angiosperms belonged to Mutator, Tc1/Mariner, and PIF/Harbinger, whereas the proportion was 35.7% in non-angiosperms. These results indicate that the drastic increase in MITE-miRNAs in angiosperms correlates with expansion of the Mutator, Tc1/Mariner, and PIF/Harbinger superfamilies.

Intrinsic MITE Features Determine Conversion to miRNAs

To identify the features of Mutator, Tc1/Mariner, and PIF/Harbinger that impact their conversion to miRNAs, we tested four parameters, including copy number of family members (which reflects their transposition activity), normalized minimal free energy (NMFE) of the MITE transcripts (which reflects stability of the hairpin structures), the guanine–cytosine (GC) content, and length of the MITE loci. We divided the Mutator, Tc1/Mariner, and PIF/Harbinger MITEs into two groups, the MITE-miRNA matching MITEs (mMITEs) and MITEs that did not match the MITE-miRNAs (nMITEs). We found that the mMITEs and nMITEs exhibited significant difference regarding the four parameters (supplementary fig. S11A–D, Supplementary Material online). Using a machine-learning approach based on a gradient boosting decision tree algorithm, we assessed relative importance of these parameters on the conversion to MITE-miRNAs. While this analysis achieved an overall accuracy of 96.8%, copy number of MITE families was the most influential feature (supplementary fig. S11E, Supplementary Material online). Further analysis revealed that proportion of the mMITEs with over 400 members was significantly higher than that of the nMITEs (supplementary fig. S11F, Supplementary Material online), suggesting that transposition activity of MITEs is pivotal for MITE-miRNA formation.

High importance scores were also retrieved for length and NMFE of MITEs (supplementary fig. S11E, Supplementary Material online), suggesting that the ability to form hairpin RNA structure is important for the generation of MITE-miRNAs. Further comparison revealed that length of MITEs has a significant impact on the miRNA repertoire of closely related species. Zea mays and Sor. bicolor of Poaceae diverged from a common ancestor about 12 My ago. Although MITEs in Z. mays (764 copies on average) were more active than those in Sor. bicolor (408 copies on average), number of MITE-miRNA families was quite close in Z. mays (21) and in Sor. bicolor (23). While NMFE of the MITEs (supplementary fig. S12, Supplementary Material online) and length distribution of MITE-miRNAs (fig. 3A) were comparable in the two species, length of MITEs in Z. mays (maximum of 132 bp) was drastically shorter than that in Sor. bicolor (maximum of 244 bp; fig. 3A). Reduction in MITE length occurred in both PIF/Harbinger (fig. 3B) and Tc1/Mariner (fig. 3C), the two dominant superfamilies in Z. mays and Sor. bicolor. Further analysis revealed that the length reduction was caused by a burst of short MITEs. In Sor. bicolor, only 15 MITEs in PIF/Harbinger and 7 MITEs in Tc1/Mariner were shorter than 100 bp. In contrast, the numbers jumped to 7,663 in Z. mays, or 4.2% of PIF/Harbinger and 18.5% of Tc1/Mariner. Thus, burst of extremely short MITEs, which likely imposes constraints on forming productive RNA hairpin structures, might have been a factor that impeded the biogenesis of MITE-miRNAs in Z. mays.

A similar pattern of length distribution was observed for the dominant Mutator superfamily in two Fabaceae species, Medicago truncatula and Glycine max (supplementary fig. S13, Supplementary Material online). While MITEs in Gl. max (1,344 copies on average) were more active than those in M. truncatula (461 copies on average), Gl. max (68) contained less MITE-miRNA families in comparison with M. truncatula (78). Again, no significant difference in NMFE was found (supplementary fig. S12, Supplementary Material online), but the length of Mutator MITEs was drastically longer in M. truncatula (maximum of 183 bp) than in Gl. max (maximum of 92 bp; supplementary fig. S13, Supplementary Material online). In conclusion, within Mutator, Tc1/Mariner, and PIF/Harbinger MITEs, copy number and length are two critical parameters determining the biogenesis of MITE-miRNAs.

A Transposition–Transcription Process for MITE-miRNA Biogenesis

To gain insight into the molecular process through which MITE-miRNAs are generated, we analyzed genomic distribution of MITE-miRNAs in the 15 examined angiosperms. We divided the genomes into four compartments based on current annotation, including exon, intron, promoter, and intergenic region. Consistent with previous reports (Rodriguez et al. 2004; Axtell et al. 2011; Moran et al. 2017; Fridrich et al. 2020), we found that the proportion of miRNAs locating in introns in the examined plants was much lower than the proportion of intronic miRNAs found in animals (supplementary table S13, Supplementary Material online). However, the proportion of MITE-miRNAs in introns was significantly higher than LTR-miRNAs, TID-miRNAs, or unclassified miRNAs (fig. 4A). Calculation of the length-normalized
proportions of MITE-miRNA and other miRNAs in different genomic compartments confirmed that MITE-miRNAs were preferentially located in introns compared with non-MITE-miRNAs (supplementary fig. S14A, Supplementary Material online). We also found that the proportion of intronic mMITEs was significantly higher than the proportion of intronic nMITEs (supplementary fig. S14B, Supplementary Material online). In the newly expanded MIR437 family in Sor. bicolor, 19 (82.6%) of the 23 members locate in introns of different host genes (fig. 4B, supplementary table S14, Supplementary Material online), suggesting that transposition in introns is the preferential route for biogenesis of this MITE-miRNA family. Using the monocot model O. sativa as an example, we found by analyzing RNA-sequencing data that the transcript levels of the MITE-miRNA host genes (in which a MITE-miRNA locates), especially those hosting the MITE-miRNAs in introns, were significantly higher than host genes of nMITEs (supplementary fig. S15, Supplementary Material online). Thus, MITE-miRNAs are more frequently located in the intron of actively transcribed host genes than non-MITE-miRNAs.

In analyzing the host genes for MITE-miRNAs, we found that many appear to be pseudogenes. This observation prompted us to quantitatively examine five species with comprehensive annotation of pseudogenes, including three grass species B. distachyon, O. sativa, and Sor. bicolor, and two legume species M. truncatula and Gl. max (Xie et al. 2019). In these species, especially M. truncatula, we found that MITE-miRNAs were indeed more frequently overlapped with pseudogenes than nMITEs (Wilcoxon signed-rank test, \( P < 0.05 \); supplementary fig. S16A, Supplementary Material online). The observation was corroborated by normalization against true protein-coding genes (supplementary fig. S16B, Supplementary Material online). Although the insertion of MITE-miRNAs may directly render the host genes to be unfunctional, preferential location of MITE-miRNAs in the introns suggests that selection of MITE-miRNAs may indirectly cause host gene pseudogenization.

Association of MITE-miRNA Target Genes with Response to Environmental Stimuli

In the examined species, except Pi. abies due to incompleteness of gene annotation (Nystedt et al. 2013), we identified 58,038 putative target genes for the 8,095 miRNAs. Consistent with previous reports that TID-miRNAs are derived from and target large gene families (Baldrich et al. 2018), we observed that the average number of target genes per TID-miRNA (10.4) was significantly higher than that of MITE-miRNA (7.5) and LTR-miRNA (5.5; supplementary fig. S17, Supplementary Material online). Gene Ontology (GO) analysis revealed that molecular functions of the MITE-miRNA target genes, compared with those of the TID-miRNAs and LTR-miRNAs, were more enriched with terms related to responses to environmental stimuli (fig. 5A, supplementary figs. S18–S20, Supplementary Material online).
One of the most frequent environmental factors associated with the MITE-miRNA target genes is temperature, with the GO terms “response to temperature stimulus,” “response to freezing,” “response to cold,” and “response to heat” being all significantly enriched (fig. 5B). We selected the MITE-miRNA Osa-miRN2285 in O. sativa (supplementary fig. S21, Supplementary Material online) that targets LOC_Os06g39750, validated by degradome sequencing data (fig. 5C and D), for analyzing its role in temperature response. LOC_Os06g39750 was a candidate gene for cold tolerance based on a quantitative trait loci analysis and exhibited a strong induction by cold-water treatment in cold tolerant rice germplasms but not in non-tolerant germplasms (Sun et al. 2018). We found that Osa-miRN2285 expression was significantly decreased after cold treatment (P < 0.01 by independent samples t-test; fig. 5E) and thus reciprocal to that of Os06g39750 in cold tolerant germplasms.

To confirm the involvement of miRN2285 in cold adaption, we employed two approaches. First, by analyzing re-sequencing data from public data set (Peng et al. 2020), we detected an SNP at the 17th position in mature Osa-miRN2285 (fig. 5D). Out of 5,459 examined germplasms, 993 possessed this SNP undermining binding to LOC_Os06g39750 and were referred to as the miRN2285 SNP type. Phenotypic records in data set (Peng et al. 2020) showed that the miRN2285SNP type germplasms exhibited significantly lower (P < 0.05 by independent samples t-test) Budding-phase cold tolerance (BPCT) than germplasms with reference miRN2285 sequence (the miRN2285 type; fig. 5F). Second, we compared the mean summer temperature at the production sites of different
Fig. 5. MITE-miRNAs facilitate response to environmental stimuli. (A) Pie charts showing the proportions of enriched GO terms associated with target genes of the MITE-miRNAs, LTR-miRNAs, and TID-miRNAs. (B) Comparison of enriched GO terms associated with response to stresses and stimuli among the target genes of different types of miRNAs. GO terms related to temperature are marked with asterisks. (C) Validation of LOC_Os06g39750 as an authentic Osa-miR2285 target by degradome sequencing. The horizontal and vertical axes represent the position of the LOC_Os06g39750 transcript and the frequency of sequenced transcript ends, respectively. Dot indicates the Osa-miRN2285-guided cleavage site. (D) Weblogo showing base complementarity between Osa-miR2285 and its target site in Os06g39750. Data were compiled from sequences of 5,459 germplasms. (E) Cold treatment suppresses Osa-miRN2285 expression. Expression of Osa-miRN2285 was compared in cold stressed and control O. sativa plants using sRNA-Seq data sets with the accession numbers GSM816687, GSM816688, GSM816689, GSM816694, GSM816695, GSM816730, GSM816731, GSM816732, GSM816739, and GSM816740. **P < 0.01 by independent samples t-test. Error bars represent the SD. (F) Comparison of BPCT between the Osa-miRN2285 type (n = 76) and Osa-miRN2285SNP type (n = 6) Oryza sativa germplasms. *P < 0.05 by independent samples t-test. The box represents values corresponding to the 25th and 75th percentile. Horizontal line inside the box represents the median. (G) Comparison of the geographical distribution of O. sativa germplasms with summer temperature in China. Geographical origins of the Osa-miRN2285 type and Osa-miRN2285SNP type germplasms are marked on the map. Distribution of the 10-year average summer temperature is superimposed on the map with temperature below 20 °C shown as blank. (H) Proportions of the Osa-miRN2285 type and Osa-miRN2285SNP type germplasms produced in areas with high (≥20 °C) and low summer temperature (<20 °C).
O. sativa germplasms in China. We found that production sites of the miRN2285 type germplasms included the northeast region that has lower summer temperature than other regions (fig. 5G). Moreover, the local summer temperature at the production sites of the 171 miRN2285SNP germplasms was significantly higher than that of the 1,177 miRN2285 type germplasms (fig. 5H, supplementary fig. S22, Supplementary Material online). Collectively, these results suggest that miRN2285-based regulation of Os06g39750 is selected in regions with lower growing season temperature where cold tolerant germplasms are predominant to facilitate the high induction of Os06g39750 after cold treatment.

**Discussion**

Biogenesis of miRNAs requires a genomic locus to be transcribed with the produced RNA capable of folding into the hairpin structure (Cui et al. 2017; Baldrich et al. 2018). Our findings provided genome-scale support for miRNA biogenesis from MITEs which meets these requirements (fig. 6). MITEs are often found close to or within genes (Santiago et al. 2002; Ohmori et al. 2008; Kuang et al. 2009), thus allowing TIR-seeded transcripts to fold into imperfect hairpin structures typical of miRNA precursors (Cui et al. 2017). We presented four lines of evidence showing that this process was highly efficient to generate novel miRNAs in angiosperm. First, we showed that MITE-miRNA families accumulated to quantities more than the TID-miRNA and LTR-miRNA families combined (fig. 2, supplementary figs. S4 and S5, Supplementary Material online). Next, in the case of the MIR437 family, which is conserved in Poaceae (Sunkar et al. 2005; Sunkar and Jagadeeswaran 2008), we discovered an expansion in Sor. bicolor due to independent transposition of MITEs belonging to the same family into different host genes (fig. 4B, supplementary figs. S6 and S7, Supplementary Material online). Finally, we found that MITE-miRNAs were often associated with host gene pseudogenization (supplementary fig. S16, Supplementary Material online), suggesting that these genes may have been selected as non-coding precursors for the miRNAs. Collectively, these findings indicate that MITEs are an important source for de novo miRNAs in angiosperms that has helped shaping the modern miRNA repertoires.

Transposable elements impose intra-genomic conflicts that can lead to an evolutionary arms race with the host genomes (Werren 2011; Luo et al. 2020). As non-autonomous elements, MITEs transpose through the action of transposases encoded by autonomous DNA transposons (Jiang et al. 2003; Yang et al. 2009), thus allowing TIR-seeded transcripts to fold into imperfect hairpin structures typical of miRNA precursors. In certain cases, selection of a miRNA might have a collateral selection of the host gene as a non-coding miRNA precursor through pseudogenization (V').
number of MITE-miRNA families and the diversity of MITE families (supplementary fig. S10, Supplementary Material online), which suggests that lineage-specific burst of MITE-miRNAs in angiosperms was due to the increasing diversity of the MITE families. A model constructed by machine learning revealed that both copy number and length of the MITEs have major effects on the biogenesis of MITE-miRNAs (fig. 3, supplementary figs. S11 and S13, Supplementary Material online). Collectively, these findings indicate that a subset of actively transcribed, intermediately sized MITEs is selectively converted to miRNAs for continuously expanding the miRNA repertoire to enhance gene regulation capacity in the host genomes (fig. 6).

In addition to the transposition–transcription model, there are other plausible genomic mechanisms for explaining the connection between MITEs and miRNAs. For instance, MITEs frequently carry non-transposon genomic sequences including gene fragments (Lisch 2002; Jiang et al. 2004; Hanada et al. 2009). It is possible that MITEs could capture an existing miRNA to generate a MITE-miRNA that duplicates the parental copy and expands the existing miRNA family. Another possibility is that the identified MITE-miRNAs might have originated from autonomous TE elements that carry a gene fragment in the TIR as in the case of the Mu4 element in maize (Lisch 2002). In this scenario, a miRNA and its homologous MITEs are derived from the same cognate autonomous partner (Piriyapongsa and Jordan 2008). Given the small number and short sequences of the miRNA loci, however, we believe the probability of generating functional miRNAs by these mechanisms might be low. Further analysis of the genomic processes associated with de novo biogenesis of miRNAs and MITEs should provide more insights into the dynamics of genome evolution.

Since their emergence, angiosperms have exceptionally radiated and surpassed all other plant lineages in both species richness and ecological dominance (Rubenstein et al. 2010; Magallón et al. 2013). Currently, angiosperms constitute the dominant vegetation on the Earth’s terrestrial surface and are indispensable to agriculture. Many genetic and genomic studies have attempted to identify the drivers of angiosperm megadiversity, which is thought to involve multivariate interactions among intrinsic traits and extrinsic forces (Vamosi and Vamosi 2011; Hausch et al. 2018; Magallón et al. 2018; Sauquet and Magallón 2018; Vamosi et al. 2018). Our findings of massive conversion of MITEs as MITE-miRNAs (fig. 6) and selection of the new gene regulators for targeting genes associated with environmental plasticity (fig. 5, supplementary figs. S18–S20, Supplementary Material online) suggest that MITE-miRNAs have the potential for driving angiosperm diversification. This notion is consistent with previous findings on the association of bursts of novel miRNAs originated from DNA transposons with the rapid diversification of vespertilionid bats (Platt et al. 2014). It is conceivable that the newly emerged miRNAs, once incorporated into the gene regulatory networks in a lineage-specific manner (Chen and Rajewsky 2007; Moran et al. 2017; Zhang et al. 2021), could lead to expression diversification of the orthologous genes which allows novel traits to evolve. Therefore, MITE-miRNAs warrant further studies as candidates for genomic innovations underlying angiosperm megadiversity.

Materials and Methods

Analysis of miRNA Conservation and Divergence

Sequences and annotations of 20,388 miRNAs in 88 species were retrieved from PmiREN1.0 (Guo et al. 2020; https://www.pmiren.com). In this data set, the miRNA loci were required to have over 75% small RNA reads corresponding to the mature and star region and <20% reads overlapping the mature and star region to eliminate possible contamination by siRNAs (Kuang et al. 2018; Guo et al. 2020). A built-in Perl script, merge_and_rank.bash, in miRDeep-P2 (Kuang et al. 2018) was used to group the miRNA families by allowing no more than two mismatches in the mature miRNA region between the members as previously defined (Meyers et al. 2008). Conservation of a miRNA family in each taxon was declared if the miRNA was detected in at least half of the examined species in that taxon. Species-specific miRNA families were defined as those were detected in only one species. The 30 taxonomic families of land plants used for pairwise comparison of miRNA families included Actinidiaceae, Amborellaceae, Araceae, Asparagaceae, Asteraceae, Brassicaceae, Bryophytes, Caricaceae, Caryophyllaceae, Convolvulaceae, Cucurbitaceae, Euphorbiaceae, Fabaceae, Fagaceae, Lythraceae, Lauraceae, Malvaceae, Musaceae, Nelumbonaceae, Oleaceae, Phrymaceae, Pinaceae, Poaceae, Rosaceae, Rutaceae, Salicaceae, Selaginellaceae, Solanaceae, Vitaceae, and Zosteraceae. Multiple in-house Perl scripts were used to compare the miRNA families among four genera and different species.

Identification of New miRNA Origination Events in Oryza and Solanum

Following the rationale showing in supplementary fig. S2, Supplementary Material online, the 428 miRNA families in O. sativa ssp. japonica were compared with those in 84 species outside of Oryza. The 379 miRNA families identified as O. sativa specific were further compared with the miRNA portfolios of seven closely related Oryza species using sRNA-Seq data. Briefly, 59 sRNA-Seq data sets (supplementary table S15, Supplementary Material online) were parsed using Trim Galore (version 0.5.0; http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/) with parameters “-length 18 -max_length 28 -small_rna -stringency 3” to remove adapters and Bowtie (1.1.2) with parameters “-v 0” to map clean reads corresponding to mature miRNAs. The ancestral states of the miRNA families were assigned following the parsimony rule previously described for identifying evolutionary events for de novo protein-coding genes (Zhang et al. 2019). The generation time of a given miRNA family was determined by the divergence time of species in which
the miRNA first appeared. The same method was employed to analyze the origination events of novel miRNA families in *Solanum* using 54 sRNA-Seq data sets from 6 species (supplementary table S16, Supplementary Material online).

### Phylogenetic Tree and Divergence Time

The phylogenetic tree of *Oryza* and *Solanum* were retrieved from the TimeTree database (Kumar et al. 2017; http://timetree.org/). TimeTree was also used to estimate the species divergence time in the two genera.

### Identification of MITE-miRNAs

Genome assemblies and annotations of the examined plants are summarized in supplementary table S4, Supplementary Material online. For these species, sequences and annotation of 6,758 miRNAs were obtained from PmiREN1.0 (Guo et al. 2020). The MITE sequences in 18 species were directly retrieved from the P-MITE Database (Chen et al. 2014; http://pmit.e.hzau.edu.cn/). MITEs in *Pi. abies* were downloaded from ConTEdb (http://genedenovowel.icb.net/81/conTEdb/index.php). MITEs in *A. trichopoda* and *Gi. biloba* were de novo identified from the genomes using MITE-hunter (Han and Wessler 2010). The outputs of MITE-hunter were manually checked for TIR and TSD sequences. Only groups with precise boundaries and TIR or TSD were considered. The consensus sequences were further considered. The consensus sequences were required to have at least ten copies in the corresponding genomes, were further screened for homology to TIRs of the MITEs and flanking TSDs or degenerated TSDs. The resultant miRNAs were defined as MITE-miRNAs. In addition, homolog search between DNA transposons from RepBase (v22.03; Bao et al. 2015) and BLAT (Kent 2002) with $E < 1 \times 10^{-10}$. The miRNAs that contain over half sequences matched to MITEs, which were required to have at least ten copies in the corresponding genomes, were further screened for homology to TIRs of the MITEs and flanking TSDs or degenerated TSDs. The resultant miRNAs were defined as MITE-miRNAs. In addition, the miRNA first appeared. The same method was employed to analyze the origination events of novel miRNA families in *Solanum* using 54 sRNA-Seq data sets from 6 species (supplementary table S16, Supplementary Material online).

### Identification of LTR-miRNAs

The LTR retrotransposon sequences in *Pi. abies*, *Z. mays*, and *Gi. biloba* were obtained from the ConTEdb (http://genedenovowel.icb.net/81/conTEdb/index.php). Plant Genome and Systems Biology Repeat Database (https://pgsb.helmholtz-muenchen.de/plant/repeatdb/index.jsp) and GigADB (http://gigadb.org/dataset/100209), respectively. Full-length LTR retrotransposons in the other 18 species were first scanned by LTR_FIND (v1.07; Xu and Wang 2007), and further filtered by LTR_retriever (v2.9.0; Ou and Jiang 2018). The resulting data set for LTR retrotransposons is stored in SOURCEFORGE (https://sourceforge.net/projects/mite-mirna/files/). To identify candidate LTR-miRNAs, BLASTN (Camacho et al. 2009) was used to search sequence similarity between miRNA precursors and the identified LTR retrotransposons. Only those with $E < 1 \times 10^{-10}$ and matched sequences longer than 50% of the precursors were considered LTR-miRNAs.

### Analysis of MITE features

Minimal free energy was calculated using RNAfold in the ViennaRNA package (version 2.2.10; Dennean 1993). NMFE was determined through normalizing MFE against the length of MITEs. The distribution of NMFE in each species was plotted using a Python (version 3.6) script. For each MITE locus, the number of MITE family members, length, and GC content were calculated using in-house Perl scripts.

### XGBoost-based Machine-Learning Model

To prepare the data sets for machine learning, four features of MITEs belonging to the *Mutator*, *PIF/Harbinger*, and *Tc1/Mariner* superfamilies in the 15 examined angiosperms were extracted as inputs. These MITEs were divided into mMITEs and nMITEs. A gradient boosting decision tree implemented in XGBoost (Extreme Gradient Boosting) was used to construct a classifying model and evaluate importance of these features. The parameters were configured by default except the learning rate (eta = 1) and maximum depth (max_depth = 2). The objective function was set as “binary:logistic,” and boosting iteration number was set as 50. After the classifying model was created, the importance scores of the four features were calculated and ranked using the “xgb.importance” function in XGBoost.

### Gene Expression Analysis in *O. sativa*

Thirteen RNA-Seq data sets for *O. sativa* ssp. *japonica* cv. Nipponbare were downloaded from the Rice Genome Annotation Project (http://rice.plantbiology.msu.edu/expression.shtml) and used for quantifying gene
expression. The data were processed using Tophat (Trappnell et al. 2009) and Cufflinks (Trapnell et al. 2010) to calculate the expression values in FPKM (mean fragments per kilobase per million mapped fragments) for either all annotated genes or genes excluding transposons (MUS Release 7.0). Host genes of MITE-miRNAs and host genes of MITEs were identified using bedtools intersect (v2.29.2; Quinlan 2014) and their expression values were extracted for comparison.

Genomic Distribution of the MIR437 Family
All MIR437 loci in four selected Poaceae species were mapped to the genomes. For loci close to or within host genes, orthologous genes were identified in other species using the Gene Orthology prediction pipeline in Ensembl Plants (Yates et al. 2022). Gene annotation was retrieved from the UniProt (Consortium 2021) and InterPro (Blum et al. 2021) databases. For MIR437 loci mapped to intergenic regions, BLASTN (Camacho et al. 2009) searches with $E < 1 \times 10^{-10}$ were implemented to identify syntenic segments against other genomes using MIR437 precursor sequences plus 500 bp flanking sequences.

Target Gene Prediction
The psRNATarget program was used to predict miRNA targets using the V2 scoring schema (Dai et al. 2018). The expectation value was set to three instead of the default setting at five to keep only the high-confidence miRNA targets.

GO Enrichment Analysis
Annotations of GO terms were downloaded from Phytozome v11 (Goodstein et al. 2011). Association of the predicted target genes for MITE-miRNAs, LTR-miRNAs, and TID-miRNAs with the GO terms was analyzed using customized scripts. Fisher's exact test with the Benjamini–Hochberg correction was used to find enriched GO terms with the adjusted $P$-value set at 0.05.

Degradome Sequencing Data Analysis
Degradome sequencing data set was downloaded from NCBI BioProject under the accession SRR039716. The CleaveLand4.0 pipeline (Addo-Quaye et al. 2009) was employed to identify miRNA cleavage sites with default parameters using O. sativa transcripts (MUS Release 7.0) as inputs.

Polymorphisms of miRN2285 and its Target Gene
Polymorphisms of Osa-miRN2285 and its target gene, LOC_Os06g39750, and phenotypic records on cold tolerance (BPCT) of different strains were retrieved from MBKBASE (Peng et al. 2020; http://mbkbase.org/). Statistical analysis was done through the independent samples $t$-test.

Association of Summer Temperature With O. sativa Germplasm Distribution
Geographical information of O. sativa germplasms in China, including longitude and latitude, was retrieved from MBKbase and shown in supplementary table S17, Supplementary Material online. Temperature data in China from 2009 to 2018 were downloaded from the National Earth System Science Data Center, National Science, and Technology Infrastructure of China (http://www.geodata.cn). The mean of summer temperature in June, July, and August were calculated by a customized MATLAB R2019b (9.7.0.1190202) script.

Identification of MITE-miRNAs and nMITEs co-localized with Pseudogenes
Annotation and genomic location of pseudogenes in B. distachyon, M. truncatula, Gl. max, O. sativa, and Sor. bicolor were obtained using the PseudoPipe workflow as previously reported (Xie et al. 2019). The obtained genomic locations were further extended by 2,000 bp to include the potential promoter regions. To obtain the genomic locations of MITEs, sequences of MITEs were aligned to reference genomes using minimap2 with default parameters (Li 2018). The relative information is provided as supplementary table S18, Supplementary Material online. Overlaps between the pseudogene loci and MITE-miRNAs and nMITEs were identified using bedtools intersect with parameter “-f 0.5.”

Other Statistical Analyses
If not stated specifically, the comparison of two distributions of values was tested with an independent samples $t$-test (two-tailed). $P$-values were shown as exact values or otherwise referenced with a symbol according to the following scales: *$P < 0.05$; **$P < 0.01$.

Acknowledgments
The authors are indebted to Hailong Chen for assistance with obtaining the temperature data. Part of the bioinformatics analysis was performed on the High-Performance Computing Platform of the Center for Life Sciences at Peking University. This work was supported by grants from the National Key Research and Development Program of China (2018YFE0204700), the National Natural Science Foundation of China (31621001 and 32070248), and Beijing Academy of Agriculture and Forestry Sciences (KJCX20200204).

Author Contributions
Conceptualization: LL, X.Y., and Z.G. Data analysis: Z.G., Z.K., and H.W. Methodology: Z.G., Z.K., H.W., and F.S.
Visualization: Z.G., Y.T., M.W., C.H., and LL. Supervision: LL. and X.Y. Writing original draft: LL, X.Y., and Z.G. Writing review and editing: LL, X.Y., Z.G., Z.K., Y.T., H.W., M.W., C.H., and F.S.

Data Availability
All codes used in this study are freely accessible via the GitHub repository (https://github.com/little-racoon/MITEmiRNA/). All MITE sequences used in this study are available in SOURCEFORGE (https://sourceforge.net/projects/mite-mirna/files/). The raw alignment results of MITE-miRNAs, LTR-miRNAs, TID-miRNAs, data set for LTR retrotransposons, information of predicted target genes, as well as raw output of CleaveLand4.0 are stored in SOURCEFORGE (https://sourceforge.net/projects/mite-mirna/files/).

References
Addo-Quaye C, Miller W, Axtell MJ. 2009. Cleaveland: a pipeline for using degradome data to find cleaved small RNA targets. Bioinformatics. 25:130–131.
Allen E, Xie Z, Gustafson AM, Sung GH, Spatafora JW, Carrington JC. 2004. Evolution of microRNA genes by inverted duplication of target gene sequences in Arabidopsis thaliana. Nat Genet. 36:1282–1290.
Axtell MJ, Meyers BC. 2018. Revisiting criteria for plant microRNA annotation in the era of big data. Plant Cell. 30:272–284.
Bao W, Kojima KK, Kohany O. 2015. Repbase update, a database of repetitive elements in eukaryotic genomes. Mob DNA. 6:11.
Bartel DP. 2004. MicroRNAs: genomics, biogenesis, mechanism, and function. Cell 116:281–297.
Bartel DP. 2018. Metazoan microRNAs. Cell 173:20–51.
Blum M, Chang HY, Chuguransky S, Grego T, Kandasaamy S, Mitchell B, Bartel DP. 2004. MicroRNAs: genomics, biogenesis, mechanism, and function. Nat Rev Genet. 5:77–89.
Fridrich A, Modepalli V, Lewandowska M, Aharoni R, Moran Y. 2020. Unravelling the developmental and functional significance of an ancient Argonaute duplication. Nat Commun. 11:6187.
Gao Z, Li J, Li L, Yang Y, Li J, Fu C, Zhu D, He H, Cai H, Li L. 2021. Structural and functional analyses of hub microRNAs in an integrated gene regulatory network of Arabidopsis. Genom Proteom Bioinf. S1672-0229(00042-00045).
Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, Mitros T, Dirks W, Bellsten U, Putnam N, et al. 2011. Phytozome: a comparative platform for green plant genomics. Nucleic Acids Res. 40:D1178–D1189.
Guo Z, Kuang Z, Wang Y, Zhao Y, Tao Y, Cheng C, Yang J, Lu X, Hao C, Wang T, et al. 2020. PmiREN: a comprehensive encyclopedia of plant miRNAs. Nucleic Acids Res. 48:D1114–D1121.
Han Y, Wessler SR. 2010. MITE-Hunter: a program for discovering miniature inverted-repeat transposable elements from genomic sequences. Nucleic Acids Res. 38:e199.
Hausch S, Vamosi SM, Fox JW. 2018. Effects of intraspecific phenotypic variation on species coexistence. Ecology. 99:1453–1462.
Jiang N, Bao Z, Zhang X, Eddy SR, Wessler SR. 2004. Pack-MULE transposable elements mediate gene evolution in plants. Nature 431:569–573.
Jiang N, Bao Z, Zhang X, Hirochika H, Eddy SR, McCouch SR, Wessler SR. 2003. An active DNA transposon family in rice. Nature 421:163–167.
Kondorosi E. 2003. The role of small RNAs in intra- and intercellular communication. Cell. 115:11–19.
Kuang Z, Wang Y, Li L, Yang X. 2018. miRDeep-P2: accurate and fast analysis of the microRNA transcriptome in plants. Bioinformatics 35:2521–2522.
Kumar S, Stecher G, Suleski M, Hedges SB. 2017. TimeTree: a resource for timelines, timetrees, and divergence times. Mol Biol Evol. 34:1812–1819.
Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics 34:3094–3100.
Lisc D. 2002. Mutator transposons. Trends Plant Sci. 7:498–504.
Lu C, Chen J, Zhang Y, Hu Q, Su W, Kuang H. 2012. Miniature inverted-repeat transposable elements (MITEs) have been accumulated through amplification bursts and play important roles in gene expression and species diversity in Oryza sativa. Mol Biol Evol. 29:1005–1017.
Luo S, Zhang H, Duan Y, Yao X, Clark AG, Lu J. 2020. The evolutionary arms race between transposable elements and piRNAs in Drosophila melanogaster. BMC Evol Biol. 20:14.

Dai X, Zhuang Z, Zhao PX. 2018. psRNATarget: a plant small RNA target analysis server (2017 release). Nucleic Acids Res. 46:W49–W54.
Denman RB. 1993. Using RNAFOLD to predict the activity of small catalytic RNAs. Biotechniques 15:1090–1095.
Dai X, Zhang H, Duan Y, Yang X, Yao X, Zhao H, Cai H, Zhang H, Lu J. 2020. The evolutionary arms race between transposable elements and piRNAs in Drosophila melanogaster. BMC Evol Biol. 20:14.
Magallón S, Hilu KW, Quandt D. 2013. Land plant evolutionary timeline: gene effects are secondary to fossil constraints in relaxed clock estimation of age and substitution rates. Ann J Bot. 100:556–573.

Magallón S, Sánchez-Reyes LÍ, Gómez-Acevedo SL. 2018. Thirty clues to the exceptional diversification of flowering plants. Ann J Bot. 105:123–151.

Meyers BC, Axtell MJ, Bartel B, Bartel DP, Baulcombe D, Bowman JL, Cao X, Carrington JC, Chen X, Green PJ, et al. 2008. Criteria for annotation of plant microRNAs. Plant Cell. 20:3186–3190.

Moran Y, Agron M, Praher D, Technau U. 2017. The evolutionary origin of plant and animal microRNAs. Nat Ecol Evol. 1:27.

Nyestedt B, Street NR, Wetterborn A, Zuccolo A, Lin YC, Scofield DG, Vezzi F, Delhomme N, Giacomello S, Alexeyenko A, et al. 2013. The Norway spruce genome sequence and conifer genome evolution. Nature 497:579–584.

Ohmori Y, Abiko M, Horibata A, Hirano HY. 2008. A transposon, Ping, is integrated into intron 4 of the DROOPING LEAF gene of rice, weakly reducing its expression and causing a mild drooping leaf phenotype. Plant Cell Physiol. 49:1176–1184.

Ou S, Jiang N. 2018. LTR_Retriever: a highly accurate and sensitive program for identification of long terminal repeat retrotransposons. Plant Physiol. 176:1410–1422.

Pearson WR, Lipman DJ. 1988. Improved tools for biological sequence comparison. Proc Natl Acad Sci USA. 85:2444–2448.

Peng H, Wang K, Cao Y, Gao Q, Li Y, Li X, Lu H, Du H, Lu M, et al. 2020. MBKbase for rice: an integrated omics knowledgebase for molecular breeding in rice. Nucleic Acids Res. 48:D1085–D1102.

Piriyapongsa J, Jordan IK. 2007. A family of human microRNA genes from miniature inverted-repeat transposable elements. PLoS One. 2:e203.

Piriyapongsa J, Jordan IK. 2008. Dual coding of siRNAs and miRNAs by plant transposable elements. RNA 14:814–821.

Platt RN II, Vandewege MW, Kern C, Schmidt CJ, Hoffmann FG, Ray DA. 2014. Large numbers of novel miRNAs originate from DNA transposons and are coincident with a large species radiation in bats. Mol Biol Evol. 31:1536–1545.

Poretti M, Praz CR, Meile L, Kálin C, Schaefer LK, Schläfli M, Widrig V, Sanchez-Vallet A, Wicker T, Bourras S. 2019. Domestication of high-copy transposons underlay the wheat small RNA response to an obligate pathogen. Mol Biol Evol. 37:839–848.

Quinlan AR. 2014. BEDTools: the Swiss-army tool for genome feature analysis. Curr Protoc Bioinformatics. 47:11.2.1-34.

Rodriguez A, Griffiths-Jones S, Ashurst JL, Bradley A. 2004. Identification of mammalian microRNA host genes and transcription units. Genome Res. 14:1902–1910.

Rodriguez RE, Schommer C, Palatnik JF. 2016. Control of cell proliferation by microRNAs in plants.Curr Opin Plant Biol. 34:68–76.

Rogers K, Chen X. 2013. Biogenesis, turnover, and mode of action of plant microRNAs. Plant Cell. 25:2383–2399.

Rubinstein CV, Girrenne P, de la Puente GS, Astini RA, Steemans P. 2010. Early Middle Ordovician evidence for land plants in Argentina (eastern Gondwana). New phytol. 188:365–369.

Santiago N, Herráiz C, Goñi JR, Meseguer X, Casacuberta JM. 2002. Genome-wide analysis of the Emigrant family of MITEs of Arabidopsis thaliana. Mol Biol Evol. 19:2285–2293.

Sauquet H, Magallón S. 2018. Key questions and challenges in angiosperm macroevolution. New phytol. 219:1170–1187.

Sun J, Yang L, Wang J, Liu H, Zheng H, Xie D, Zhang M, Feng M, Jia Y, Zhao H, et al. 2018. Identification of a cold-tolerant locus in rice (Oryza sativa L) using bulked segregant analysis with a next-generation sequencing strategy. Rice 11:24.

Sunkar R, Githe T, Jain PK, Zhu JK. 2005. Cloning and characterization of microRNAs from rice. Plant Cell. 17:1397–1411.

Sunkar R, Jagadeeswaran G. 2008. In silico identification of conserved microRNAs in large number of diverse plant species. BMC Plant Biol. 8:37.

Trapnell C, Pachter L, Salzberg SL. 2009. Tophat: discovering splice junctions with RNA-seq. Bioinformatics 25:i1105–i1111.

Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L. 2010. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol. 28:511–515.

Vamosi JC, Magallón S, Mayrose I, Otto SP, Sauquet H. 2018. Macroevolutionary patterns of flowering plant speciation and extinction. Annu Rev Plant Biol. 69:685–706.

Vamosi JC, Vamosi SM. 2011. Factors influencing diversification in angiosperms: at the crossroads of intrinsic and extrinsic traits. Am J Bot. 98:460–471.

Voinnet O. 2009. Origin, biogenesis, and activity of plant microRNAs. Curr Opin Plant Biol. 13:669–687.

Wenk HJ. 2011. Selfish genetic elements, genetic conflict, and evolutionary innovation. Proc Natl Acad Sci USA. 108:10863–10870.

Xie J, Li Y, Liu Y, Zhao Y, Li B, Ingvason PK, Zhang D. 2019. Evolutionary origins of pseudogenes and their association with regulatory sequences in plants. Plant Cell. 31:563–578.

Xu Z, Wang H. 2007. LTRINDER: an efficient tool for the prediction of full-length LTR retrotransposons. Nucleic Acids Res. 35:265–268.

Yang G, Dong J, Chandrasekharan M, Hall T. 2001. Kiddo, a new transposable element family closely associated with rice genes. Mol Genet Genomics. 266:417–424.

Yang G, Nagel DH, Feschotte C, Hancock CN, Wessler SR. 2009. Tuned for transposition: molecular determinants underlying the hyperactivity of a Stowaway MITE. Science. 325:1391–1394.

Yang X, Zhang H, Li L. 2012. Alternative mRNA processing increases the complexity of microRNA-based gene regulation in Arabidopsis. Plant J. 70:421–431.

Yates AD, Allen J, Amode RM, Azov AG, Barba M, Becerrra A, Bhai J, Campbell L, Carbajo Martinez M, Chakravilli M, et al. 2022. Ensemble Genomes 2022: an expanding genome resource for non-vertibrates. Nucleic Acids Res. 50:D996–d1003.

Zhang H, Guo Z, Zhuang Y, Suo Y, Du J, Gao Z, Pan Ji, Li L, Wang T, Xiao L, et al. 2021. MicroRNA775 regulates intrinsic leaf size and reduces cell wall pectin levels by targeting a galactosyltransferase gene in Arabidopsis. Plant Cell. 33:581–602.

Zhang L, Ren Y, Yang T, Li G, Chen J, Gschwend AR, Yu Y, Hou G, Zi J, Zhou R, et al. 2019. Rapid evolution of protein diversity by de novo origination in Oryza. Nat Ecol Evol. 3:679–690.