

RESEARCH

Open Access



Amplification of miniature inverted-repeat transposable elements and the associated impact on gene regulation and alternative splicing in mulberry (*Morus notabilis*)

Youchao Xin, Bi Ma, Zhonghuai Xiang and Ningjia He*

Abstract

Background: Miniature inverted-repeat transposable elements (MITEs) are common in eukaryotic genomes, and are important for genomic evolution.

Results: In the present study, the identification of MITEs in the mulberry genome revealed 286,122 MITE-related sequences, including 90,789 full-length elements. The amplification of mulberry MITEs and the influence of MITEs on the evolution of the mulberry genome were analyzed. The timing of MITE amplifications varied considerably among the various MITE families. Fifty-one MITE families have undergone a single round of amplification, while the other families developed from multiple amplifications. Most mulberry MITEs were inserted near genes and some could regulate gene expression through small RNAs. An analysis of transgenic plants indicated that MITE insertions can upregulate the expression of a target gene. Moreover, MITEs are frequently associated with alternative splicing events (exonizations).

Conclusion: The data presented herein provide insights into the generation of MITEs as well as their impact on gene regulation and evolution in mulberry.

Keywords: MITEs, Amplification, Gene expression, Small RNA, Alternative splicing

Background

Miniature inverted-repeat transposable elements (MITEs) are very short, deletion derivatives of autonomous DNA transposons [1, 2]. They were originally discovered in the maize genome, and are widespread among animals and plants [3, 4]. MITEs and autonomous DNA transposons share common characteristics, including the presence of a terminal inverted-repeat (TIR) flanked by a target site duplication (TSD). The TIR and TSD suggest that most MITEs are derived from autonomous DNA elements, including *Tc1/Mariner* elements [5], *PIF/Harbinger* [1, 6, 7], *hAT* [8, 9], and *Mutator* [10]. Although MITEs lack an open reading frame encoding a transposase, their transposition is mediated by transposases associated with autonomous DNA transposons [11, 12].

In plants, MITEs often cover a considerable portion of the genome, including up to 10% of the *Oryza sativa* genome and 8% of the *Medicago truncatula* genome [13]. In rice, active MITEs have been detected (e.g., *mPing* and *mGing*) [11, 14], and the *Stowaway* MITE has been developed as a genetic engineering tool for transferring heterologous genes [15]. MITEs are often transcribed with plant genes [16, 17], and can influence genomic evolution and gene expression [18–22]. Specifically, MITEs usually downregulate gene expression [17, 23, 24], but MITEs containing regulatory motifs can have the opposite effect [20, 22]. Furthermore, MITEs can encode small RNAs that regulate the expression of target genes at the transcriptional or post-transcriptional levels [17, 25]. The structural similarity between MITEs and microRNA genes suggests that MITE-derived small RNAs may be generated via the microRNA pathway [26]. However, the MITE-derived small RNAs in Solanaceae species are most likely

* Correspondence: hejia@swu.edu.cn

State Key Laboratory of Silkworm Genome Biology, Southwest University, Beibei, Chongqing 400715, China



generated by the small interfering RNA (siRNA) biogenesis pathway [17]. Additionally, small RNAs derived from MITEs are important for silencing transposable elements (TEs) through stem-loop structures [27].

Alternative splicing is a common post-transcriptional regulatory process that increases transcriptome and proteome diversity in eukaryotic organisms. Alternative splicing is reportedly important for development [28, 29] and stress responses [30, 31]. Analyses of RNA sequences have revealed abundant alternatively spliced, intron-containing transcripts in *Arabidopsis thaliana* (61%), *Oryza sativa* (33%), *Zea mays* (40%), and *Glycine max* (63%) [32–35]. Moreover, alternative splicing based on TE activities has been confirmed in previous studies. For example, more than 5% of the alternative splicing occurring in humans is associated with *Alu* elements [36]. Additionally, in *A. thaliana*, more than half of the expressed *Ty1/Copia* elements are spliced [37]. Another study proved that *Alu* elements help modulate alternative splicing [38]. However, it remains unclear whether MITEs can mediate alternative splicing in plants.

Mulberry (*Morus* sp.) is a well-known food source for silkworms (*Bombyx mori* L.) and is an economically, ecologically, and medically important plant species. *Morus notabilis* has a relatively small genome (approximately 357 Mb), which has been sequenced [39]. Its genomic data may be useful for thoroughly characterizing mulberry MITEs. In this study, we identified the MITEs in the mulberry genome and analyzed their amplification patterns, effects on gene regulation, and evolution.

Results

Identification of MITEs in the mulberry genome

A total of 286,122 MITE-related sequences were detected in the mulberry genome, including 90,789 (31.73%) full-length elements (Table 1). The sequences of all MITE families and their distribution in mulberry genomes are available from the *Morus notabilis* transposable element database (<http://morus.swu.edu.cn/mntedb/>). The MITE-related sequences covered 13.83% of the mulberry genome, which was more than the corresponding coverage of the *O. sativa* genome (10% [13]).

On the basis of sequence similarities, the mulberry MITEs were classified into 232 families. The TIR and TSD sequences were used to categorize the mulberry MITE families into the following four superfamilies: *Tc1/Mariner* (5.11%), *PIF/Harbinger* (4.67%), *Mutator* (0.44%), and *hAT* (0.97%) (Table 1). The ratio of full-length to partial MITEs in individual superfamilies varied from 13 to 60%, with only *Tc1/Mariner* having a value exceeding 50% (Table 1).

Amplification of MITE families in the mulberry genome

We investigated the amplification of 195 MITE families. Pairwise sequence diversities were calculated and histograms were drawn for the full-length MITE sequences from each family. A total of 100 families exhibited a multimodal distribution, while 51 and 44 exhibited unimodal and bimodal distributions, respectively (Fig. 1a and b). The wave histograms of the pairwise diversities suggested that each family underwent rapid amplification during evolution [40].

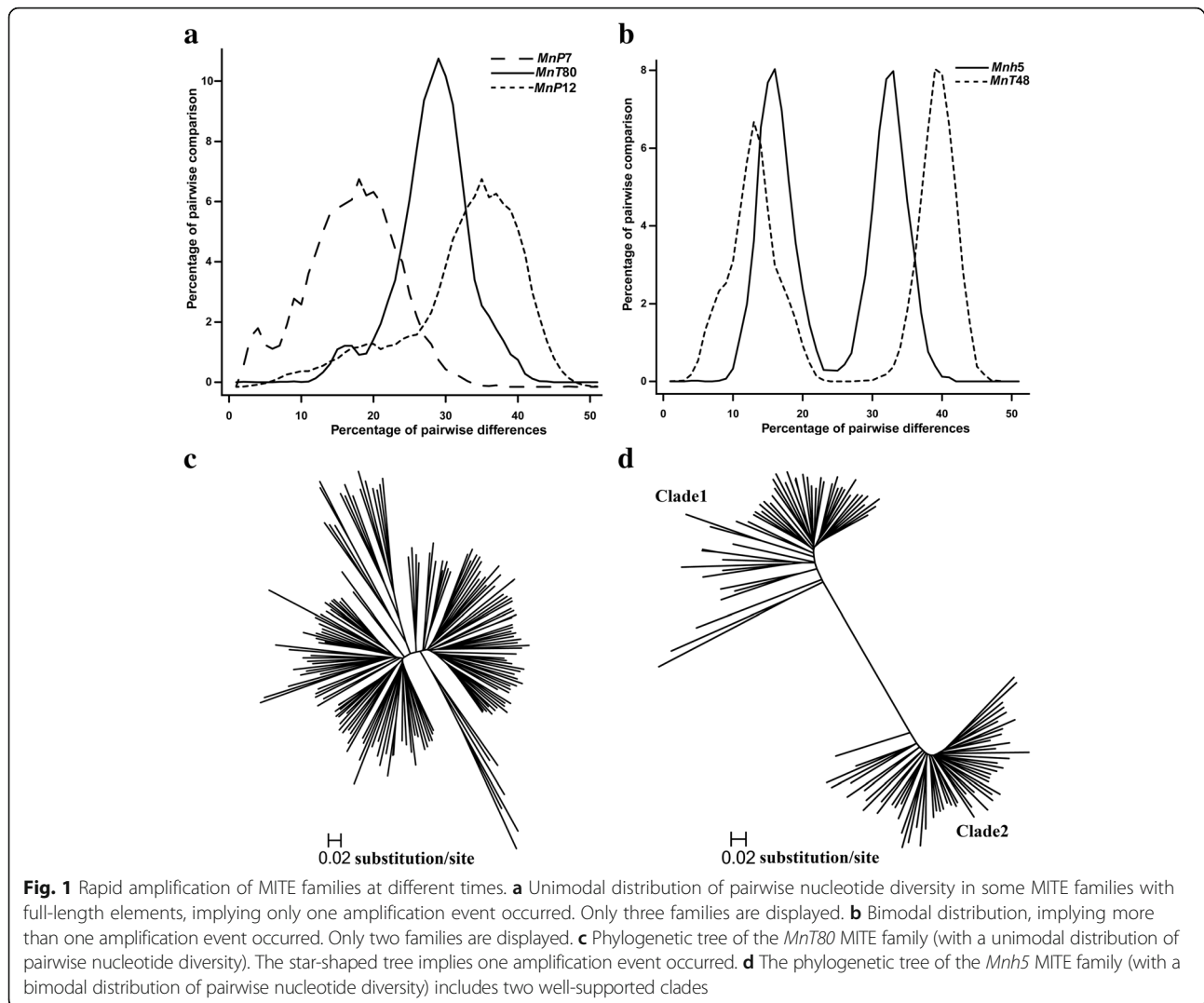
The histograms for these MITE families included unimodal peaks at different diversity levels, suggesting that the amplification of individual families occurred at distinct time points (Fig. 1a). The amplification time for *MnP7* (average pairwise nucleotide diversity of 0.160) was estimated to be 12.3 million years ago. In contrast, the amplification time for *MnP12* (average pairwise nucleotide diversity of 0.312) was estimated to be 24 million years ago [41].

Three phylogenetic trees were constructed for the families that produced unimodal and bimodal peaks. The MITE families with a unimodal peak distribution generated star-shaped phylogenetic trees, implying that their amplification was rapid and originated from a single master element (Fig. 1c). In contrast, the MITE families with a bimodal peak distribution were divided into two clades, implying that they had multiple ancestors or experienced multiple amplifications (Fig. 1d). Thus, some MITE families experienced one amplification event, while other families underwent multiple amplification events during evolution.

Table 1 Summary of the MITE superfamilies identified in *Morus notabilis*

Superfamily	Family Number	Total Elements	Full-length Number	Partial-length Number	Full-length/Partial-length (%)	Percentage of Genome (%)	TSD
<i>Tc1/Mariner</i>	89	116,427	43,474	72,953	59.59	5.11	TA
<i>PIF/Harbinger</i>	74	89,616	15,560	74,056	21.01	4.67	TWA
<i>hAT</i>	21	10,288	1570	8718	18.01	0.97	8 bp
<i>Mutator</i>	8	8130	921	7209	12.78	0.44	9 bp
<i>DTx</i>	40	61,661	29,264	32,397	90.33	2.63	
Total	232	286,122	90,789	195,333	46.48	13.83	

DTx represents an unknown superfamily



To further explore MITE amplification, we studied the insertional polymorphism of MITEs in various mulberry species. Four polymorphic MITE loci were randomly chosen for PCR amplification using primers designed against the sequences flanking the MITEs (Fig. 2). The following three banding patterns were observed: (1) a band for the expected full site; (2) a band for the expected empty site; or (3) no band (i.e., no amplification). PCR amplification may have failed because the primers did not anneal to the expected sequence owing to a mutation in this region. We detected one locus that had MITEs in all analyzed samples (Fig. 2a and c), suggesting that this MITE insertion was fixed in *Morus* species. In some cases (Fig. 2b and d), the results clearly indicated that MITEs may have been activated following a polyploidization event. An analysis of the sequences of the extracted PCR bands for the expected full and empty sites confirmed that the difference between the

sequences corresponding to the upper and lower bands in the gel was the presence or absence of a MITE (Fig. 3).

Localization of MITEs in the mulberry genome

The MITEs inserted in gene sequence (GS) and intergenic sequence (IS) were counted, and the data were used to construct a regression curve for each family (Fig. 4a; $r^2 = 0.65$, $P < 0.05$). Four MITE families had more elements in IS regions than expected (Fig. 4a), suggesting that these four MITE families were excluded from GS regions during evolution, while the other MITE families exhibited a linear and uniform distribution of elements. Moreover, for each superfamily, the distribution of MITEs in gene structure regions, as well as the 2000-bp sequences upstream of the start codon and downstream of the stop codon, were analyzed to evaluate whether the MITE insertion sites were preferentially close to genes (Fig. 4b). With the exception of the *Tc1*/

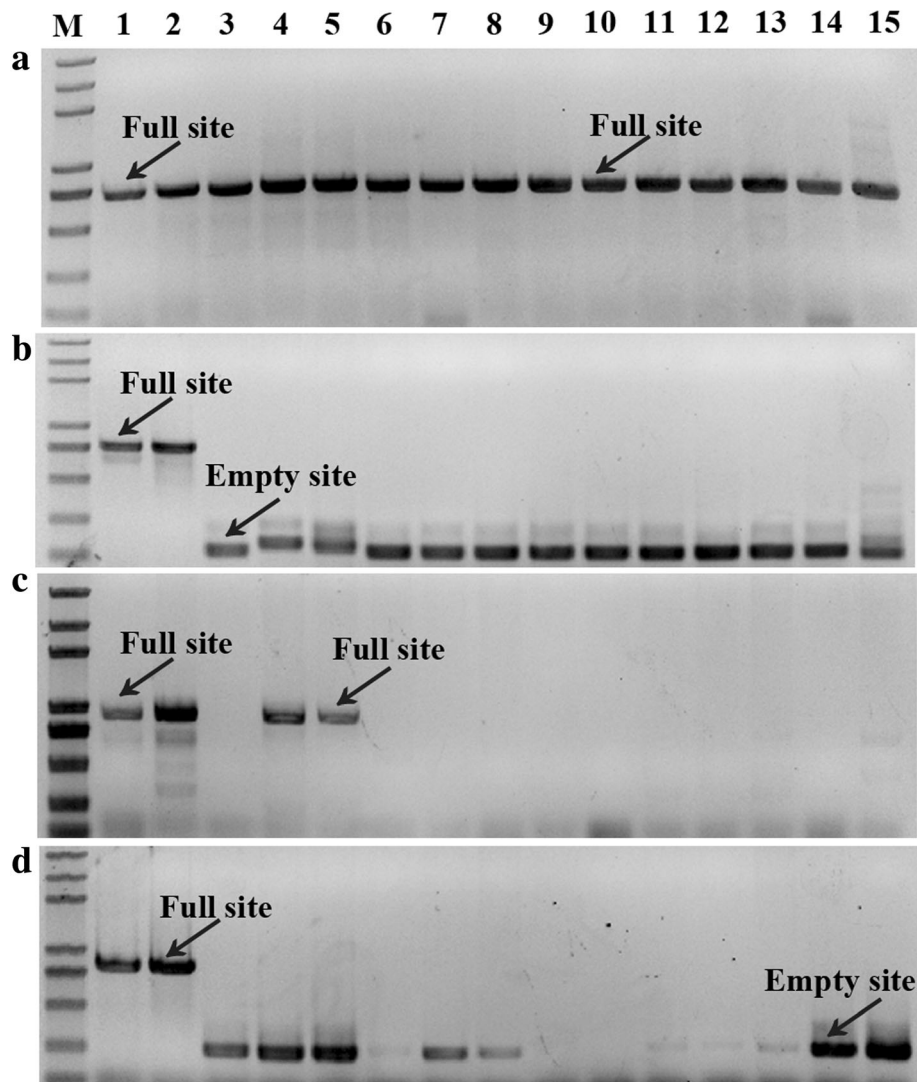


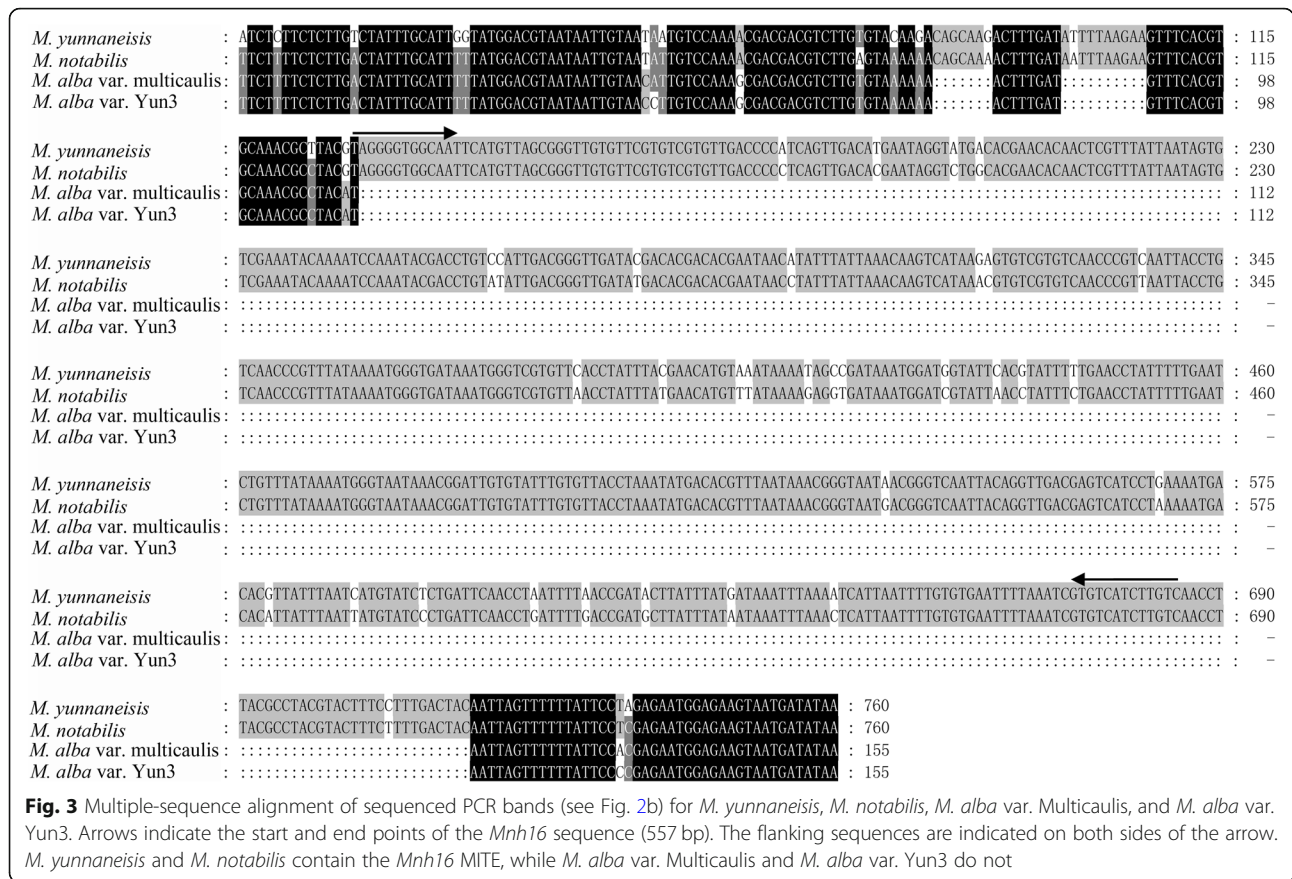
Fig. 2 Site-specific PCR analysis. **a** *Mnh16_scaffold1108_216257–216,737*, **b** *Mnh16_scaffold1960_162356–162,912*, **c** *Mnh16_scaffold93_462750–463,388*, and **d** *Mnh16_scaffold897_369175–369,745* in the following mulberry resources: 1, *M. notabilis*; 2, *M. yunnaneisis*; 3, *M. alba* var. Yun3; 4, *M. mongolica* Schneid.; 5, *M. wittiorum* Hand.-Mazz.; 6, *M. alba* var. Jinqiang63; 7, *M. alba* var. Taiwandaguo; 8, *M. alba* var. Xinjiaposijiguo; 9, *M. alba* var. Zhenzhubai; 10, *M. alba* var. Lunjiao109; 11, *M. alba* var. Multicaulis; 12, *M. alba* var. Hongguo1; 13, *M. alba* var. Hongguo2; 14, *M. alba* var. Da10; and 15, *M. nigra*. M, marker. Bands corresponding to a full site or empty site are indicated

Mariner superfamily, all MITE superfamilies tended to be inserted near genes.

After filtering out genes with a low RPKM value (i.e., < 1), the remaining genes with MITE insertion in exons were identified and analyzed. Among the five MITE superfamilies, the *PIF/Harbinger* MITEs had the highest expression ratio (i.e., the ratio of the total number of expressed MITEs to the total number of genes) (Fig. 4c). The proportions of MITE-associated genes that were expressed in five tissues were as follows: flower (1.65%), bud (1.58%), leaf (1.58%), bark (1.53%) and root (1.40%). These results implied that the MITE-associated gene expression ratios were relatively consistent among the five analyzed tissues.

The *MnM2* MITE induced the ectopic expression of *MnANR*

In the current study, we examined the *MnM2* MITE inserted near the gene encoding the phosphate-responsive 1 family protein as well as the sequence polymorphisms in this region in different mulberry resources (Fig. 5a). The effect of *MnM2* on gene expression was subsequently investigated. Three expression vectors (Fig. 5b) were constructed and inserted into tobacco leaves via *A. tumefaciens*-mediated transformation. The expression levels of *MnANR*, which can change the color of tobacco flowers from red to white [42], in T_0 transgenic tobacco seedlings were determined in a qRT-PCR assay. The expression levels in the transgenic tobacco seedlings



were significantly higher than in the wild type. Additionally, transgenic seedlings carrying *MnM2* had significantly higher *MnANR* expression levels than those lacking *MnM2* (Fig. 5c). The activity of *MnANR* in the seedlings with *MnM2* was higher than that in the seedlings without *MnM2*, which led to a deeper change of tobacco color from red to white (see Additional file 2: Figure S1).

Small RNAs derived from mulberry MITEs

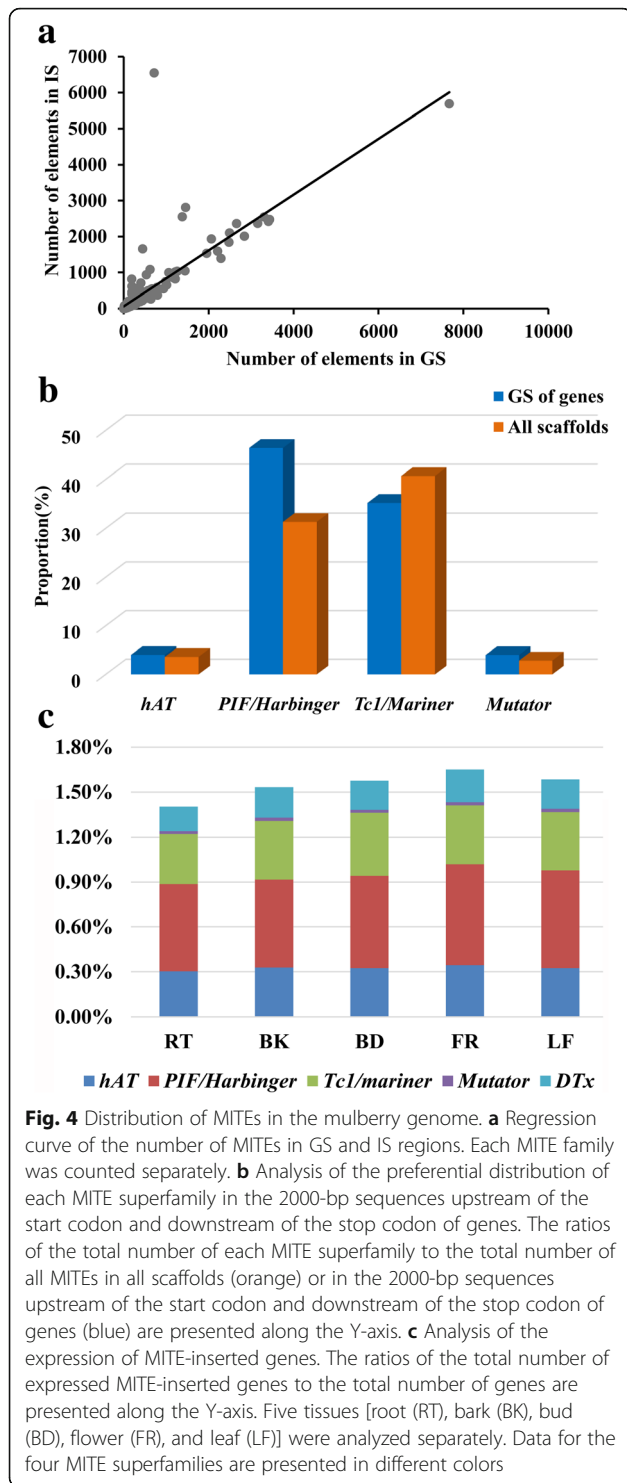
In mulberry, 45,577 (15.9%) MITE sequences completely matched small RNA sequences. We analyzed the ratio of the number of small RNA-containing MITEs to the total number of MITEs in each superfamily. The proportions of MITEs in the four superfamilies were as follows: *PIF/Harbinger* (20.68%), *hAT* (20.13%), *Tc1/Mariner* (6.40%) and *Mutator* (3.43%) (Table 2). The *PIF/Harbinger* superfamily had the highest ratio of the number of small RNA-containing MITEs to the total number of MITEs. Additionally, there was no correlation between the ratio and the number of MITEs in each superfamily. Interestingly, 64.7% of the MITE-derived small RNAs were produced by MITEs located close to a gene (see Additional file 2: Figure S2). The mulberry MITE-derived small RNAs were 23–25 nt long, but 24-nt small RNAs were

dominant (Fig. 6a). These observations were consistent with the reported results for Solanaceae species [17].

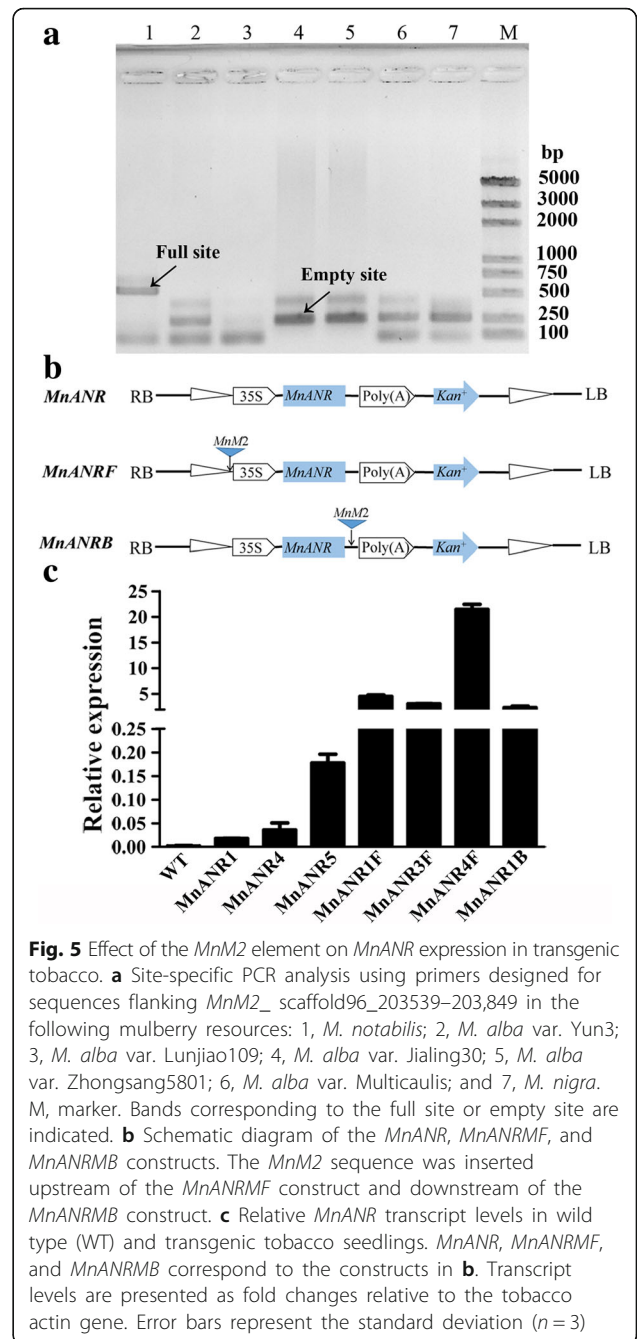
The relative positions of small RNAs in MITEs were used to determine which part of the MITE sequences generated small RNAs because MITEs vary in length. For all MITE families, the parts of the full-length MITE sequences that could generate small RNAs were investigated. We observed that small RNAs with complete coverage were distributed throughout the MITEs, with one obvious peak in the central region and two at the termini (Fig. 6b). To investigate the variations in different MITE families, elements with nearly identical lengths from 13 MITE families were independently analyzed. Surprisingly, we detected considerable variations in the positions of small RNAs in different MITE families. Four of the 13 MITE families generated small RNAs mainly from the central region (Fig. 6c), while six other families generated small RNAs predominantly from the termini (Fig. 6d).

Alternative splicing of mulberry genes related to MITEs

Parts of MITEs can be retained in mature mRNAs via splicing (exonization), which is facilitated by sequence motifs that resemble splice sites (see Additional file 2: Figure S3 for a model of exonization). We comprehensively



surveyed the association between MITEs and four basic modes of alternative splicing in five tissues. Of the four modes, alternative 5' and 3' splice sites were the predominant MITE-related modes (based on the ratio of the number of MITE-related alternative splicing events to the number of all alternative splicing events) (Fig. 7a). The



proportion of MITE-related alternative splicing in five tissues was as follows: flower (5.00%), leaf (4.26%), root (4.02%), bark (3.31%) and bud (2.58%), (Fig. 7a); thus, MITEs had more exonization in flowers. Moreover, we analyzed the ratio of the number of MITEs at alternative splicing sites to the number of all alternative splicing events (Fig. 7b). The proportion of MITEs at alternative splicing sites in the five tissues was as follows: flower (0.51%), bark (0.50%), bud (0.48%), root (0.45%) and leaf (0.43%) (Fig. 7b), interestingly, the MITE ratio at the alternative splicing site was consistent across the five tissues

Table 2 Associations between MITEs and mulberry genes

Superfamily	Total Elements	Associated with Genes	Expressed with Genes	Related with small RNAs	Related with small RNAs/Total Elements (%)
<i>Tc1/Mariner</i>	116,427	34,418	743	7451	6.40
<i>PIF/Harbinger</i>	89,616	45,392	430	18,534	20.68
<i>hAT</i>	10,288	3924	204	2071	20.13
<i>Mutator</i>	8130	3897	13	279	3.43
Total	224,461	87,631	1390	28,335	12.62

compared to the MITE ratio at the exonization. To further verify the MITE-related alternative splicing, we analyzed a mulberry pathogenesis-related protein PR-4 gene (*MnPR-4*) in three mulberry resources (Fig. 8). The intron of this gene includes a MITE, which results in the second exon being spliced (Fig. 8a). Through PCR analysis, we found that *MnPR-4* was polymorphic in different mulberry resources (Fig. 8b). Through reverse transcription PCR analysis, we found that *MnPR-4* was missing from some mulberry resources, and that *MnPR-4* alternative splicing was also missing (Fig. 8c). These PCR products were verified by cloning and sequencing.

Discussion

Detection and characterization of mulberry MITEs

A previous study suggested that the number of MITEs is associated with genome size [13]. For example,

papaya has a relatively small genome (342.68 Mb) and only one MITE family comprising 538 MITE-related sequences. Conversely, apple has a larger genome (881.28 Mb) and contains 180 MITE families with 237,302 MITE-related sequences [13]. In this study, however, we identified 232 MITE families with 286,122 MITE-related sequences in the *M. notabilis* genome (357 Mb) (Table 1). Although the mulberry genome is similar in size to that of papaya, it contains considerably more MITE sequences, likely because the mulberry MITEs underwent multiple amplifications during evolution. One amplification event was responsible for 26% of the MITEs in the mulberry genome. Additionally, *Tc1/Mariner* is the most abundant of the four known MITE superfamilies in mulberry, likely because of its proportion of full-length elements, which can be relatively easily activated.

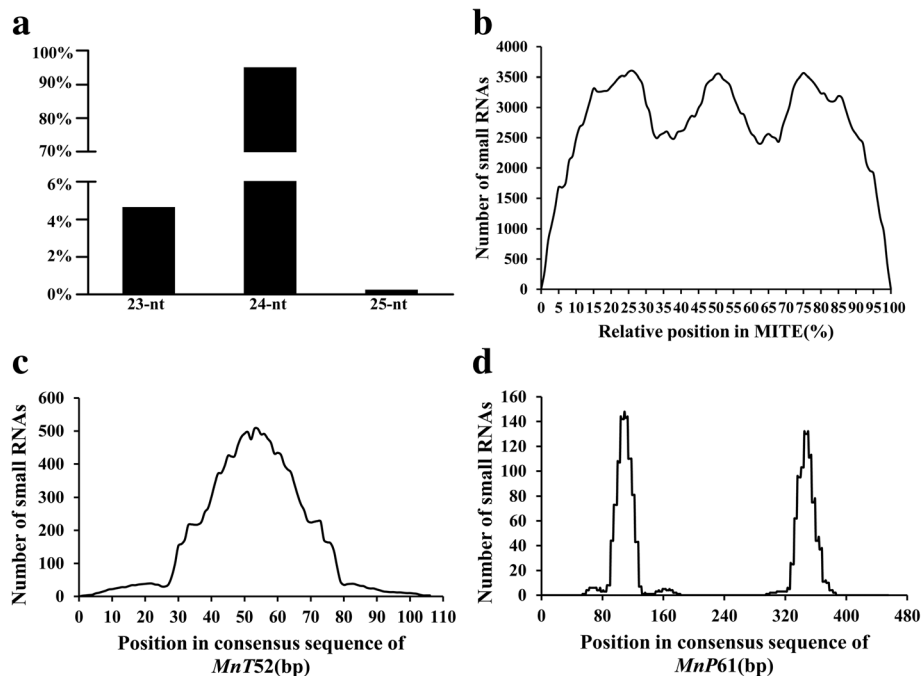


Fig. 6 Length distributions and positions of MITE-derived small RNAs. **a** Length distribution of MITE-derived small RNAs. The ratios of the total number of small RNAs (for each length) to the total number of all small RNAs are presented along the Y-axis. **b** The relative positions of MITE-derived small RNAs in all full-length MITEs. **c** Small RNAs are predominantly generated from the central regions of MITEs in four of the 13 MITE families investigated. Only the *MnT52* family is presented. **d** Small RNAs are predominantly generated from the terminal regions of MITEs in six of the 13 MITE families investigated. Only the *MnP61* family is presented

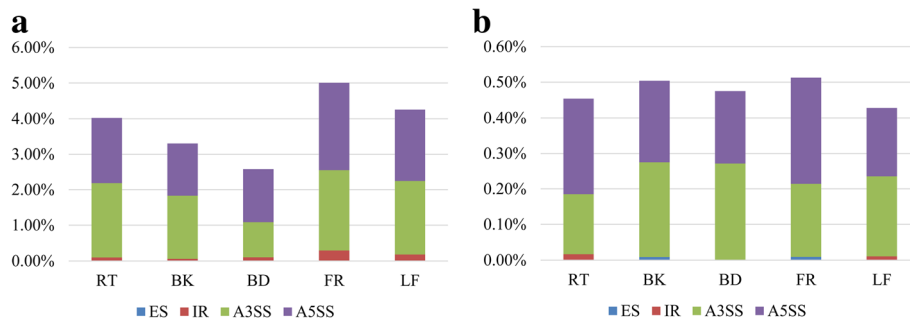


Fig. 7 Analysis of alternatively spliced genes associated with MITEs. **a** The ratios of the total number of MITE-associated alternatively spliced genes to the total number of all alternatively spliced genes are presented along the Y-axis. **b** The ratios of the number of MITEs at alternative splicing sites to the number of all alternatively spliced genes are presented along the Y-axis. Five tissues [root (RT), bark (BK), bud (BD), flower (FR), and leaf (LF)] were analyzed separately. The four basic modes of alternative splicing [exon skipping (ES), intron retention (IR), alternative 3' splice site (A3SS), and alternative 5' splice site (A5SS)] are indicated in different colors

Intermittent generation of MITEs

The members of each MITE family are similar in sequence and length, which suggests the families underwent at least one amplification during evolution [43]. This suggestion was verified by histograms of pairwise nucleotide diversity among mulberry MITEs (Fig. 1). Analyses of different clades implied that many MITE families have undergone several rounds of amplification. The diversity in the amplification times among MITE families indicates these amplifications were not due to genome-wide duplications. It is possible that MITE amplifications in the *M. notabilis* genome occurred only sporadically. The MITEs may have been activated by a “genomic shock” or the temporary activation of a cognate transposase [44]. Indeed, in rice, *mPing* may have been activated by irradiation, cell culture, or domestication [11, 45, 46]. Most MITEs in the *M. notabilis*

genome are relatively old and may be the result of selection or genetic drift during the long evolutionary history of the species.

Impact of MITEs on mulberry gene expression

Our analysis suggested that *M. notabilis* MITEs are widely distributed in the genome and have been preferentially inserted upstream and downstream of genes (< 2000 bp), which is similar to what has been reported for other species such as *O. sativa* [47]. Although many MITEs are associated with genes, only a few studies have surveyed the effects of MITEs on gene expression [17, 20, 22–24]. In rice, the expression of the *Ghd2* gene is suppressed by a MITE in the 3' untranslated region [24]. In maize, the variability at the *ZmNAC111* locus is likely caused by an 82-bp MITE inserted in the gene promoter, which may repress gene expression via the

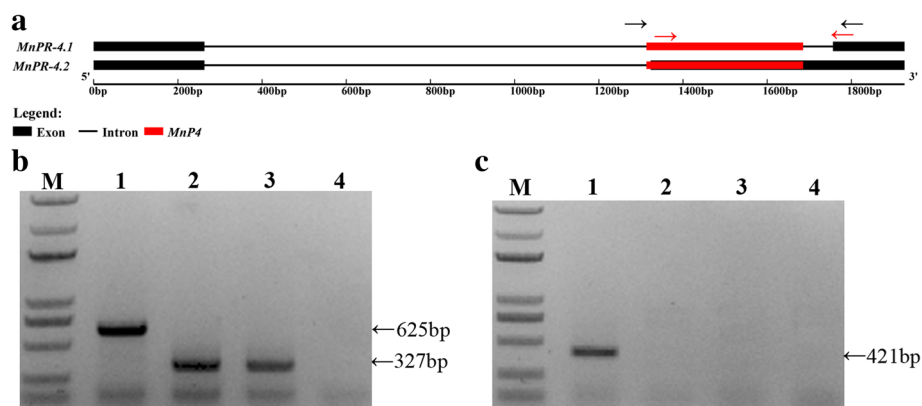


Fig. 8 Alternative splicing of the *MnPR-4* gene related to MITEs in different mulberry resources. **a** Genomic structure of MITE-related *MnPR-4* splice variants. The arrowheads indicate primer positions for PCR amplification. **b** Site-specific PCR analysis using primers designed for the sequences flanking *MnPR-4_scaffold205_783088–783,460*. The primer positions are shown with the black arrows in **a**. **c** Detection of alternatively spliced transcripts. The reverse transcription PCR analysis of *MnPR-4* expression in the leaf. The primer positions are shown with red arrows in **a**. The PCR products were separated by agarose gel electrophoresis, and the resulting gel was stained with ethidium bromide. PCR product sizes are indicated on the right. 1–4 represent *M. notabilis*; *M. alba* var. Taiwan; *M. alba* var. Hongguo1 and water. M, marker

RNA-directed DNA methylation pathway [23]. In Solanaceae species, MITEs generate small RNAs and regulate gene expression through the small RNA silencing pathway [17]. A previous analysis of promoter activity revealed that MITE *kiddo* was responsible for approximately 20% of the expression levels of a neighboring gene in both transiently and stably transformed rice calli [22]. Moreover, when the methylation of *kiddo* was blocked with 5-azaC, the accumulation of *ubiquitin2* transcripts reportedly increased 3-fold [22]. These results indicate *kiddo* has dual functions that regulate gene expression. In this study, we detected transcriptional differences depending on the presence of MITEs in transgenic tobacco. Specifically, MITEs can influence the expression levels of genes with which they are associated. Thus, MITEs may have an effect on the evolution of gene expression. Furthermore, MITEs may be useful for regulating the expression of transgenes.

A thorough characterization of the impact of MITEs on mulberry gene expression requires a systemic investigation of MITE-derived small RNAs. The transcribed MITEs themselves may form double-stranded RNAs. Plant siRNAs include the 21-nt and 24-nt classes. A previous study determined that 21-nt siRNAs regulate mRNAs post-transcriptionally, while 24-nt siRNAs suppress gene expression at the transcriptional level via RNA-dependent DNA methylation and heterochromatin maintenance [48]. Our analyses uncovered 45,577 (15.9%) MITE sequences that completely matched a small RNA sequence. The MITE-derived small RNAs in the *M. notabilis* genome are mainly (95%) 24 nt long, similar to those in Solanaceae species [17]. Interestingly, 64.7% of the MITE-derived small RNAs were produced by MITEs located close to a gene, possibly because these MITEs may be more likely to be transcribed than those in intergenic regions. When considering all MITEs, small RNAs appear to be derived slightly more frequently from the central region and termini than from other regions (Fig. 6). However, the positions of small RNAs on MITEs vary considerably among the MITE families. The members of some MITE families produce small RNAs mainly from the central region, with very few from the termini, while the opposite trend occurs in other families. These differences are likely mediated by specific mechanisms that will need to be elucidated in future studies.

The evolution of the mulberry genome was accelerated by MITEs

Our results revealed that many mulberry MITEs are located in alternative splice sites (Fig. 7). These observations imply that MITEs are important for alternative splicing. In mulberry, some MITEs are located in alternative splice sites and many MITEs are associated with

alternative splicing. Moreover, most MITE-containing exons are alternatively spliced. This phenomenon suggests that alternative splicing caused by transposons may contribute to genetic disease. For example, an *Alu* insertion can reportedly cause a genetic disorder in humans [49, 50]. Transposable elements are a major contributor to the generation of lineage-specific exons in primates [51]. The insertion of TEs in intronic regions of genes may lead to mutations that create new exons [51, 52]. The exonization of TEs occurred frequently during human evolution. Moreover, noncoding RNAs, TEs, and alternative splicing are all involved in regulating gene expression. On the basis of our analyses of the mulberry genome, we determined that many noncoding RNAs are derived from MITEs. In human cells, 5S-OT regulates the *in trans* alternative splicing of multiple genes via TE/anti-TE pairings with target genes. MITEs can also form sense/antisense transcripts. We hypothesize that the insertion of MITEs in intronic regions of genes creates new exons, and such MITEs, some of which are alternatively spliced, were major contributors to the development of species-specific exons in mulberry. Additionally, these MITEs help modulate alternative splicing events. Furthermore, the exonization of MITEs can generate noncoding RNAs that may regulate gene expression levels. These processes may have enabled the evolution of mulberry species.

Conclusions

Although many plant MITEs have been investigated, there have been few relevant studies on mulberry. We herein revealed the MITE amplification patterns in the mulberry genome. Our analyses of transgenic plants and small RNAs suggest that MITE insertions may regulate mulberry gene expression in diverse ways (e.g., epigenetic modifications and the production of new regulatory motifs). The frequent association of MITEs with alternative splicing and activation likely contributed to the evolution of mulberry. Our results provide insights into the generation of MITEs and as their contribution to the genetic regulation and evolution of mulberry.

Methods

Genome-wide identification of mulberry MITEs

Mulberry genomic sequences were downloaded from the MorusDB database (<https://morus.swu.edu.cn/morusdb/>). The MITE-Hunter, MITE Digger, and Repetitive Sequence with Precise Boundaries programs were used to identify potential MITEs in the mulberry genome with default parameters [53–55]. Candidate MITEs were detected based on the 80–80–80 rule to remove redundancies and were grouped into MITE families [56]. To mine all copies of each MITE family member, representative sequences with the expected length and structure

for each family were used as queries to clarify the distribution and locations of MITEs in the mulberry genome with the RepeatMasker v3.2.9 program (<http://www.repeatmasker.org/>) and the Cross_match search engine.

All MITE families were classified into superfamilies based on similarities in the TSD and TIR sequences. Each MITE family was designated *MnX#*, where *Mn*, *X*, and *#* represent *M. notabilis*, the superfamily, and the family serial number, respectively. For the superfamily, *T*, *h*, *P*, *M*, and *N* correspond to *Tc1/Mariner*, *hAT*, *PIF/Harbinger*, *Mutator*, and unknown, respectively.

Analysis of the amplification of MITE families

To assess the amplification of each MITE family, the MUSCLE v3.8 program was used to align the full-length sequences of the elements [57]. MITE sequences that were up to 3-bp short at their termini relative to the conserved element were considered to be full-length sequences. Neighbor-joining trees (pairwise deletion for gaps and the Kimura two-parameter substitution model) were constructed for each MITE family with the MEGA 6 program [58]. Pairwise nucleotide diversity was defined as the number of mismatches divided by the alignment length. Every gap was considered as a single mismatch. A Perl script was used to calculate and visualize the pairwise nucleotide diversity among MITE family members. A substitution rate of 1.3×10^{-8} base substitutions per site per year was used to estimate the divergence time between two sequences [41].

Survey of mulberry genomic variations caused by MITE insertions

All the primers used for PCR and qRT-PCR are listed in Additional file 1: Table S1. Genomic variations caused by MITE insertions were checked by PCR. Primers were designed based on the sequences flanking MITEs. The PCR amplifications were completed with DNA samples from the following 17 mulberry resources (*M. notabilis*, $2n = 2x = 14$; *Morus yunnaneisis*, $2n = 2x = 14$; *Morus alba* var. Yun3, $2n = 4x = 28$; *Morus mongolica* Schneid., $2n = 5x = 35$; *Morus wittiorum* Hand.-Mazz., $2n = 7x = 49$; *M. alba* var. Jinqiang63, $2n = 4x = 28$; *M. alba* var. Taiwandaguo, $2n = 4x = 28$; *M. alba* var. Xinjiaposijiguo, $2n = 4x = 28$; *M. alba* var. Zhenzhubai, $2n = 4x = 28$; *M. alba* var. Lunjiao109, $2n = 4x = 28$; *M. alba* var. Jialing30, $2n = 8x = 56$; *M. alba* var. Zhongsang5801, $2n = 4x = 28$; *M. alba* var. Multicaulis, $2n = 4x = 28$; *M. alba* var. Hongguo1, $2n = 4x = 28$; *M. alba* var. Hongguo2, $2n = 4x = 28$; *M. alba* var. Da10, $2n = 5x = 35$; and *Morus nigra*, $2n = 44x = 308$). Each 20 μ L reaction contained 20 ng genomic DNA, 2 mM PCR buffer, 0.2 mM each primer, 0.2 mM each dNTP, and 1 unit Taq polymerase with 2.5 mM $MgCl_2$ (Takara Biotechnology Company, Dalian, China). The PCR amplification protocol was as follows:

94 °C for 4 min; 32 cycles of 94 °C for 30 s, 58 °C for 30 s, and 72 °C for 1 min; 72 °C for 7 min. The amplification products were sequenced and compared to further analyze MITE insertions.

Mulberry genes associated with MITEs

Files with the coordinates of the predicted genes in scaffolds were obtained from the MorusDB database [59]. To determine whether MITEs were preferentially associated with predicted genes, all sequences were divided into two parts, namely the GS and IS. The GS included the region from 2000-bp sequences upstream of the start codon to 2000-bp downstream of the stop codon of genes (if the distance between two adjacent genes is less than 2000 bp, the spaced sequences were directly used in analyses). All other sequences were considered part of the IS. Additionally, a Perl script was written to identify MITE-associated genes. To test whether MITE superfamilies were preferentially associated with specific genomic regions, we determined the proportions of each superfamily in GS regions or in the scaffold.

The expression levels of MITE-inserted genes were analyzed in root, bark, bud, flower, and leaf tissues. Genes with MITEs inserted in their exons and an RPKM value (i.e., reads per kilobase of exon per million mapped reads) ≥ 1 , were included in our analyses. The transcriptome data for the five analyzed tissues were downloaded from the MorusDB database.

Vector construction, plant transformation, RNA isolation, and quantitative RT-PCR

The *M. notabilis* MITE sequences inserted near a gene were amplified by PCR using specific primers and then cloned into the *Bst*EII and *Hind*III or *Sal*I and *Eco*RI restriction enzyme sites of the pLGNL vector (upstream of the 35S promoter or downstream of the target gene). The target gene encoded an anthocyanidin reductase (*MnANR*, GenBank accession no. EXB31407.1), which can change the color of tobacco flowers from red to white [42], and was cloned into the *Sal*I and *Eco*RI or *Kpn*I and *Bam*HI restriction enzyme sites of the pLGNL vector. The correct orientation of the inserted amplified fragments was confirmed by DNA sequencing. The two recombinant vectors were introduced into *Agrobacterium tumefaciens* strain GV3101 cells, which were then co-cultivated with tobacco leaf sections as previously described [60].

Total RNA was isolated from the leaves of tobacco or mulberry plants with the TRIzol reagent (Invitrogen, USA). The extracted RNA was used as the template for synthesizing the first-strand cDNA with the PrimeScript™ RT-PCR Kit (Takara, Dalian, China). A quantitative real-time (qRT)-PCR assay was completed with the QuantiNova™ SYBR Green PCR Kit (Qiagen, Valencia,

CA, USA) and the StepOnePlus™ Real-time PCR system (Applied Biosystems, USA). The *ACTIN* gene was used as an internal control for normalizing target gene expression levels.

Identification of MITE-derived small RNAs

To identify MITE-derived small RNAs, mulberry leaf small RNA data were downloaded from the NCBI Sequence Read Archive database (project number: SRP032829; <https://www.ncbi.nlm.nih.gov/sra/>). All mulberry small RNA sequences were used as queries for BLAST searches for MITE sequences. Small RNA sequences that completely matched a MITE sequence were considered to be derived from the MITE [61]. These MITE-derived small RNAs were then mapped to the full-length MITE sequences to determine their positions in the MITEs. Because of the variability in MITE lengths, the small RNAs were mapped based on their relative positions in MITE sequences. For example, if a small RNA completely matched a 100-bp full-length MITE between nucleotides 30 and 50, the small RNA was mapped to the 30–50% region of this MITE. The number of small RNAs at each relative position was calculated and visualized. The number of small RNAs at each relative position was also investigated for individual MITE families. To limit biases, only families containing MITEs with similar lengths and more than 100 small RNAs were analyzed.

Alternatively spliced mulberry genes associated with MITEs

Four basic modes of alternative splicing were analyzed (i.e., exon skipping, alternative 5' splice site, alternative 3' splice site, and intron retention). Alternatively spliced genes were considered to be associated with MITEs if the alternatively spliced sites were on MITEs or if MITEs were only found in alternative splicing variants of the genes. The analyses involved root, bark, bud, flower, and leaf tissues, and the transcriptome data for these five tissues were downloaded from the MorusDB database.

Additional files

Additional file 1: Table S1. Primers used for PCR or qRT-PCR amplification of specific genes or transposable elements. (DOCX 14 kb)

Additional file 2: Figure S1. Tobacco flowers of wild-type (WT) and transgenic lines. Transgenic lines were transformed with *MnANR4*, *MnANR5*, *MnANR1F*, or *MnANR4F*. **Figure S2.** Analysis of the distribution of MITE-derived small RNAs in IS and GS regions. The ratios of the total number of MITE-derived small RNAs in IS or GS regions to the total number of MITE-derived small RNAs are presented along the Y-axis. **Figure S3.** Schematic model of MITE exonization. A MITE is inserted into the intron of a gene. During evolution, mutations within pseudo-splice sites activate the MITE insertion sites (marked by black arrows), and part of the MITE sequence is recognized as a new exon ('exonized'). (ZIP 756 kb)

Abbreviations

GS: Gene sequence; IS: Intergenic sequence; MITEs: Miniature inverted-repeat transposable elements; qRT-PCR: Quantitative real-time polymerase chain reaction; TIR: Terminal inverted-repeat; TSD: Target site duplication

Authors' contributions

YX, BM, ZX, and NH conceived and designed the experiments; YX conducted the experiments; YX and BM analyzed the data; YX wrote the manuscript and NH revised the paper. All authors read and approved the final manuscript.

Funding

This project was funded by the National Key Research and Development Program (No. 2018YFD1000602), Chongqing Research Program of Basic Research and Frontier Technology (cstc2018jcyjAX0407), Fundamental Research Funds for the Central Universities (SW18040), Natural Science Foundation of China (No. 31572323), the special fund for Agro-scientific research in the public interest of China (No. 201403064), and the "111" Project (B12006).

Availability of data and materials

Mulberry genomic and transcriptome sequences were downloaded from the MorusDB database. The mulberry leaf small RNA data were downloaded from the NCBI Sequence Read Archive database under the project number: SRP032829.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Received: 15 April 2019 Accepted: 17 June 2019

Published online: 25 June 2019

References

- Feschotte C, Mouches C. Evidence that a family of miniature inverted-repeat transposable elements (MITEs) from the Arabidopsis thaliana genome has arisen from a pogo-like DNA transposon. *Mol Biol Evol.* 2000;175:730–7.
- Zhang Q, Arbuckle J, Wessler SR. Recent, extensive, and preferential insertion of members of the miniature inverted-repeat transposable element family heartbreaker into genic regions of maize. *Proc Natl Acad Sci U S A.* 2000;973:1160–5.
- Bureau TE, Wessler SR. Stowaway: a new family of inverted repeat elements associated with the genes of both monocotyledonous and dicotyledonous plants. *Plant Cell.* 1994;66:907–16.
- Tu Z. Three novel families of miniature inverted-repeat transposable elements are associated with genes of the yellow fever mosquito, *Aedes aegypti*. *Proc Natl Acad Sci U S A.* 1997;9414:7475–80.
- Feschotte C, Swamy L, Wessler SR. Genome-wide analysis of mariner-like transposable elements in rice reveals complex relationships with stowaway miniature inverted repeat transposable elements (MITEs). *Genetics.* 2003; 1632:747–58.
- Zhang XY, Jiang N, Feschotte C, Wessler SR. PIF- and pong-like transposable elements: distribution, evolution and relationship with tourist-like miniature inverted-repeat transposable elements. *Genetics.* 2004;1662:971–86.
- Zerjal T, Joets J, Alix K, Grandbastien MA, Tenaillon MI. Contrasting evolutionary patterns and target specificities among three tourist-like MITE families in the maize genome. *Plant Mol Biol.* 2009;711(2):99–114.
- Moreno-Vazquez S, Ning JC, Meyers BC. hATpin, a family of MITE-like hAT mobile elements conserved in diverse plant species that forms highly stable secondary structures. *Plant Mol Biol.* 2005;586:869–86.
- Depra M, Ludwig A, Valente VL, Loreto EL. Mar, a MITE family of hAT transposons in drosophila. *Mob DNA.* 2012;31:13.
- Yang GJ, Hall TC. MDM-1 and MDM-2: two mutator-derived MITE families in rice. *J Mol Evol.* 2003;563:255–64.
- Jiang N, Bao ZR, Zhang XY, Hirochika H, Eddy SR, McCouch SR, Wessler SR. An active DNA transposon family in rice. *Nature.* 2003;4216919:163–7.

12. Yang GJ, Nagel DH, Feschotte C, Hancock CN, Wessler SR. Tuned for transposition: molecular determinants underlying the hyperactivity of a stowaway MITE. *Science*. 2009;325:946:1391–4.
13. Chen J, Hu Q, Zhang Y, Lu C, Kuang H. P-MITE: a database for plant miniature inverted-repeat transposable elements. *Nucleic Acids Res*. 2014; 42(Database issue):D1176–81.
14. Dong HT, Zhang L, Zheng KL, Yao HG, Chen J, Yu FC, Yu XX, Mao BZ, Zhao D, Yao J, et al. A gajjin-like miniature inverted repeat transposable element is mobilized in rice during cell differentiation. *BMC Genomics*. 2012;13:135.
15. Fattash I, Bhardwaj P, Hui C, Yang G. A rice stowaway MITE for gene transfer in yeast. *PLoS One*. 2013;8:5:e64135.
16. Oki N, Yano K, Okumoto Y, Tsukiyama T, Teraishi M, Tanisaka T. A genome-wide view of miniature inverted-repeat transposable elements (MITEs) in rice, *Oryza sativa ssp. japonica*. *Genes & genetic systems*, vol. 834; 2008. p. 321–9.
17. Kuang H, Padmanabhan C, Li F, Kamei A, Bhaskar PB, Ouyang S, Jiang J, Buell CR, Baker B. Identification of miniature inverted-repeat transposable elements (MITEs) and biogenesis of their siRNAs in the Solanaceae: new functional implications for MITEs. *Genome Res*. 2009;19:42–56.
18. Santiago N, Herraiz C, Goni JR, Messeguer X, Casacuberta JM. Genome-wide analysis of the emergent family of MITEs of *Arabidopsis thaliana*. *Mol Biol Evol*. 2002;19:2:2285–93.
19. Sampath P, Lee SC, Lee J, Izzah NK, Choi BS, Jin M, Park BS, Yang TJ. Characterization of a new high copy stowaway family MITE, BRAMI-1 in Brassica genome. *BMC Plant Biol*. 2013;13:56.
20. Naito K, Zhang F, Tsukiyama T, Saito H, Hancock CN, Richardson AO, Okumoto Y, Tanisaka T, Wessler SR. Unexpected consequences of a sudden and massive transposon amplification on rice gene expression. *Nature*. 2009;461:7267:1130–4.
21. El Amrani A, Marie L, Ainouche A, Nicolas J, Coue I. Genome-wide distribution and potential regulatory functions of AtATE, a novel family of miniature inverted-repeat transposable elements in *Arabidopsis thaliana*. *Mol Genet Genomics*. 2002;267:459–71.
22. Yang GJ, Lee YH, Jiang YM, Shi XY, Kertbundit S, Hall TC. A two-edged role for the transposable element kiddo in the rice ubiquitin2 promoter. *Plant Cell*. 2005;17:5:1559–68.
23. Mao HD, Wang HW, Liu SX, Li Z, Yang XH, Yan JB, Li JS, Tran LSP, Qin F. A transposable element in a NAC gene is associated with drought tolerance in maize seedlings. *Nat Commun*. 2015;6:8326.
24. Shen JQ, Liu JH, Xie KB, Xing F, Xiong F, Xiao JH, Li XH, Xiong LZ. Translational repression by a miniature inverted-repeat transposable element in the 3' untranslated region. *Nat Commun*. 2017;8:14651.
25. Cai YM, Zhou Q, Yu CX, Wang XM, Hu SN, Yu J, Yu XM. Transposable-element associated small RNAs in *Bombyx mori* genome. *PLoS One*. 2012; 7(5):e36599.
26. Piriyaopongsa J, Marino-Ramirez L, Jordan IK. Origin and evolution of human microRNAs from transposable elements. *Genetics*. 2007;176:2:1323–37.
27. McCue AD, Slotkin RK. Transposable element small RNAs as regulators of gene expression. *Trends Genet*. 2012;28:12:616–23.
28. Ruhl C, Stauffer E, Kahles A, Wagner G, Drechsel G, Ratsch G, Wachter A. Polypyrimidine tract binding protein homologs from *Arabidopsis* are key regulators of alternative splicing with implications in fundamental developmental processes. *Plant Cell*. 2012;24:11:4360–75.
29. Staiger D, Brown JWS. Alternative splicing at the intersection of biological timing, development, and stress responses. *Plant Cell*. 2013;25:10:3640–56.
30. Li WF, Lin WD, Ray P, Lan P, Schmidt W. Genome-wide detection of condition-sensitive alternative splicing in *Arabidopsis* roots. *Plant Physiol*. 2013;162:3:1750–63.
31. Cui P, Zhang S, Ding F, Ali S, Xiong L. Dynamic regulation of genome-wide pre-mRNA splicing and stress tolerance by the Sm-like protein LSm5 in *Arabidopsis*. *Genome Biol*. 2014;15:1:R1.
32. Marquez Y, Brown JW, Simpson C, Barta A, Kalyna M. Transcriptome survey reveals increased complexity of the alternative splicing landscape in *Arabidopsis*. *Genome Res*. 2012;22:6:1184–95.
33. Zhang GJ, Guo GW, Hu XD, Zhang Y, Li QY, Li RQ, Zhuang RH, Lu ZK, He ZQ, Fang XD, et al. Deep RNA sequencing at single base-pair resolution reveals high complexity of the rice transcriptome. *Genome Res*. 2010;20:5: 646–54.
34. Shen Y, Zhou Z, Wang Z, Li W, Fang C, Wu M, Ma Y, Liu T, Kong LA, Peng DL, et al. Global dissection of alternative splicing in paleopolyploid soybean. *Plant Cell*. 2014;26:3:996–1008.
35. Thatcher SR, Zhou WG, Leonard A, Wang BB, Beatty M, Zastrow-Hayes G, Zhao XY, Baumgarten A, Li BL. Genome-wide analysis of alternative splicing in *Zea mays*: landscape and genetic regulation. *Plant Cell*. 2014;26:9:3472–87.
36. Sorek R, Ast G, Graur D. Alu-containing exons are alternatively spliced. *Genome Res*. 2002;12:7:1060–7.
37. Oberlin S, Sarazin A, Chevalier C, Voinnet O, Mari-Ordóñez A. A genome-wide transcriptome and translome analysis of *Arabidopsis* transposons identifies a unique and conserved genome expression strategy for Ty1/Copia retroelements. *Genome Res*. 2017;27:9:1549–62.
38. Hu SS, Wang XL, Shan G. Insertion of an Alu element in a lncRNA leads to primate-specific modulation of alternative splicing. *Nat Struct Mol Biol*. 2016; 23:11:1011–9.
39. He NJ, Zhang C, Qi XW, Zhao SC, Tao Y, Yang GJ, Lee TH, Wang XY, Cai QL, Li D, et al. Draft genome sequence of the mulberry tree *Morus notabilis*. *Nat Commun*. 2013;4:2445.
40. Rogers AR, Harpending H. Population-growth makes waves in the distribution of pairwise genetic-differences. *Mol Biol Evol*. 1992;9:3:552–69.
41. Ma J, Bennetzen JL. Rapid recent growth and divergence of rice nuclear genomes. *Proc Natl Acad Sci U S A*. 2004;101:34:12404–10.
42. Han YP, Vimolmangkang S, Soria-Guerra RE, Korban SS. Introduction of apple ANR genes into tobacco inhibits expression of both CHI and DFR genes in flowers, leading to loss of anthocyanin. *J Exp Bot*. 2012;63:7:2437–47.
43. Han MJ, Shen YH, Xu MS, Liang HY, Zhang HH, Zhang Z. Identification and evolution of the silkworm helitrons and their contribution to transcripts. *DNA Res*. 2013;20:5:471–84.
44. McClintock B. The significance of responses of the genome to challenge. *Science*. 1984;226:4676:792–801.
45. Nakazaki T, Okumoto Y, Horibata A, Yamahira S, Teraishi M, Nishida H, Inoue H, Tanisaka T. Mobilization of a transposon in the rice genome. *Nature*. 2003;421:6919:170–2.
46. Naito K, Cho E, Yang G, Campbell MA, Yano K, Okumoto Y, Tanisaka T, Wessler SR. Dramatic amplification of a rice transposable element during recent domestication. *Proc Natl Acad Sci U S A*. 2006;103:47:17620–5.
47. Jiang N, Feschotte C, Zhang XY, Wessler SR. Using rice to understand the origin and amplification of miniature inverted repeat transposable elements (MITEs). *Curr Opin Plant Biol*. 2004;7:2:115–9.
48. Baulcombe D. RNA silencing in plants. *Nature*. 2004;431:7006:356–63.
49. Knebelmann B, Forestier L, Drouot L, Quinones S, Chuet C, Benassy F, Saus J, Antignac C. Splice-mediated insertion of an Alu sequence in the COL4A3 mRNA causing autosomal recessive Alport syndrome. *Hum Mol Genet*. 1995;4:4:675–9.
50. Vervoort R, Gitzelmann R, Lissens W, Liebaers I. A mutation (IVS8-0.6kbpdelTC) creating a new donor splice site activates a cryptic exon in an Alu-element in intron 8 of the human beta-glucuronidase gene. *Hum Genet*. 1998;103:6:686–93.
51. Sorek R. The birth of new exons: mechanisms and evolutionary consequences. *Rna*. 2007;13:10:1603–8.
52. Lev-Maor G, Sorek R, Shomron N, Ast G. The birth of an alternatively spliced exon: 3' splice-site selection in Alu exons. *Science*. 2003;300:5623:1288–91.
53. Han YJ, Wessler SR. MITE-hunter: a program for discovering miniature inverted-repeat transposable elements from genomic sequences. *Nucleic Acids Res*. 2010;38(22):e199.
54. Yang GJ. MITE digger, an efficient and accurate algorithm for genome wide discovery of miniature inverted repeat transposable elements. *BMC Bioinformatics*. 2013;14:186.
55. Lu C, Chen JJ, Zhang Y, Hu Q, Su WQ, Kuang HH. 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*. 2012;29:3:1005–17.
56. Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capi P, Chalchoub B, Flavell A, Leroy P, Morgante M, Panaud O, et al. A unified classification system for eukaryotic transposable elements. *Nat Rev Genet*. 2007;8:12:973–82.
57. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res*. 2004;32:5:1792–7.
58. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: Molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol*. 2013;30:12:2725–9.
59. Li T, Qi X, Zeng Q, Xiang Z, He N. MorusDB: a resource for mulberry genomics and genome biology. *Database*. 2014;2014:bau054.
60. Klee H, Horsch R, Rogers S. Agrobacterium-mediated plant transformation and its further applications to plant biology. *Annu Rev Plant Phys*. 1987;38: 467–86.

61. Zhou M, Tao G, Pi P, Zhu Y, Bai Y, Meng X. Genome-wide characterization and evolution analysis of miniature inverted-repeat transposable elements (MITEs) in moso bamboo (*Phyllostachys heterocycla*). *Planta*. 2016;244:775–87.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

