Dual coding of siRNAs and miRNAs by plant transposable elements

JITTIMA PIRIYAPONGSA and I. KING JORDAN
School of Biology, Georgia Institute of Technology, Atlanta, Georgia 30332-0230, USA

ABSTRACT

We recently proposed a specific model whereby miRNAs encoded from short nonautonomous DNA-type TEs known as MITEs evolved from corresponding ancestral full-length (autonomous) elements that originally encoded short interfering (siRNAs). Our miRNA-origins model predicts that evolutionary intermediates may exist as TEs that encode both siRNAs and miRNAs, and we analyzed *Arabidopsis thaliana* and *Oryza sativa* (rice) genomic sequence and expression data to test this prediction. We found a number of examples of individual plant TE insertions that encode both siRNAs and miRNAs. We show evidence that these dual coding TEs can be expressed as readthrough transcripts from the intronic regions of spliced RNA messages. These TE transcripts can fold to form the hairpin (stem–loop) structures characteristic of miRNA genes along with longer double-stranded RNA regions that typically are processed as siRNAs. Taken together with a recent study showing Drosha independent processing of miRNAs from Drosophila introns, our results indicate that ancestral miRNAs could have evolved from TEs prior to the full elaboration of the miRNA biogenesis pathway. Later, as the specific miRNA biogenesis pathway evolved, and numerous other expressed inverted repeat regions came to be recognized by the miRNA processing endonucleases, the host gene-related regulatory functions of miRNAs emerged. In this way, host genomes were afforded an additional level of regulatory complexity as a by-product of TE defense mechanisms. The siRNA-to-miRNA evolutionary transition is representative of a number of other regulatory mechanisms that evolved to silence TEs and were later co-opted to serve as regulators of host gene expression.

Keywords: transposable elements; miRNA; siRNA; arabidopsis; rice

INTRODUCTION

The phenomenon of RNA-mediated gene regulation was originally discovered in plants (Matzke and Matzke 2004). Plant biologists found that post-transcriptional gene silencing (PTGS) seemed to involve RNA or DNA sequence interactions between transgenes, or transgenes and homologous plant genes, which led to sequence-specific RNA degradation (Napoli et al. 1990; van der Krol et al. 1990; de Carvalho et al. 1992; van Blokland et al. 1994). It soon became apparent that plant RNA viruses could also stimulate PTGS. Transgenic tobacco plants that expressed a truncated form of a viral coat gene recovered from initial infection with the virus and ultimately became resistant (Lindbo et al. 1993). This resistance was found to be conferred through degradation of viral RNA. Subsequently, PTGS was shown to serve as a natural mechanism employed by plants to defend against viral infection (Covey et al. 1997; Ratcliff et al. 1997). Ultimately, these findings led to the notion that a number of plant gene silencing mechanisms initially evolved as defense mechanisms against invading genetic elements (Matzke et al. 2000).

The broader significance of RNA-mediated gene regulation became widely apparent only later, when the specific role of double-stranded RNA (dsRNA) in RNA interference (RNAi) was elucidated for *Caenorhabditis elegans* (Fire et al. 1998). RNAi in *C. elegans* was related to genome defense mechanisms by studies showing that RNAi deficient mutants lost the ability to silence Tc*1* transposable elements (TEs) in the germline (Ketting et al. 1999; Tabara et al. 1999). The mechanism behind RNAi-based silencing of *C. elegans* TEs was found to be based on the production of dsRNAs from the terminal inverted repeat (TIR) sequences found at the ends of Tc*1* elements (Sijen and Plasterk 2003). This work demonstrated that RNAi is initiated by read-through transcription of full-length Tc*1* elements,

Reprint requests to: I. King Jordan, School of Biology, Georgia Institute of Technology, 310 Ferst Drive, Atlanta, GA 30332-0230, USA; e-mail: king.jordan@biology.gatech.edu; fax: (404) 894-0519.

Article published online ahead of print. Article and publication date are at http://www.rnajournal.org/cgi/doi/10.1261/rna.916708.

which then fold into "snap-back" structures with the complementary sequences of the TIRs bound as dsRNA (Fig. 1). These dsRNA TIR sequences are processed by the RNAi enzymatic machinery to yield short interfering RNAs (siRNAs) that silence expression via mRNA degradation of the transposase gene required for Tc*1* transposition. The sequence specificity of the mRNA degradation is caused by binding of the TIR-derived single-stranded siRNAs to complementary sequences of the transposase encoding mRNA. Later, TE-encoded siRNAs were shown to silence the highly active *MuDR* TE family in maize (Slotkin et al. 2005). In light of the ability to defend against viral infection and TE mobilization, RNAi has been considered as an immune system for the genome (Plasterk 2002).

As described above, the connection between the siRNA molecules that mediate RNA-based gene silencing and TEs, or viruses, has been appreciated since PTGS and RNAi were first studied. MicroRNAs (miRNAs) are a related class of short RNA molecules with an analogous functional role in RNAi (Ambros 2004; Bartel 2004). miRNAs are processed from the dsRNA regions of short, ~70–90 base pairs (bp), stem–loop (hairpin) RNA structures by the same endonuclease, Dicer (or Dicer-like in plants), which cleaves siRNAs from longer dsRNA sequences. A connection between TEs and miRNAs was more recently established when a number of miRNA genes were found to be derived from TE sequences (Mette et al. 2002; Smalheiser and Torvik 2005; Borchert et al. 2006; Piriyapongsa and Jordan 2007; Piriyapongsa et al. 2007).

In the human genome, a group of related miRNA genes was found to be derived from the Made1 family of TEs (Piriyapongsa and Jordan 2007). Made1 elements (Morgan 1995; Oosumi et al. 1995; Smit and Riggs 1996) are members of a specific class on DNA-type TEs known as miniature inverted-repeat transposable elements (MITEs) (Bureau and Wessler 1992, 1994). MITEs are short non-autonomous derivatives of full-length DNA-type elements (Feschotte and Mouches 2000; Feschotte et al. 2002). Full-length DNA-type elements are, typically, several kilobases in length and contain a single open reading frame, which encodes the transposase enzyme that catalyzes transposition, flanked by two TIR sequences on either end of the elements (Fig. 1A). As is the case with the Tc*1* elements of *C. elegans*, full-length transcripts of DNA-type elements can fold into snap-back structures with the two TIRs forming a dsRNA region (Fig. 1B). This dsRNA region can be processed to yield siRNAs that silence expression of the elements. MITEs are shorter sequences of ~80–500 bp, which lack the internal ORF of full-length elements but retain the TIRs (Fig. 1C). So MITEs are closer to being palindromes, and read-through transcription of MITEs will lead to RNA sequences that can fold into hairpin structures reminiscent of the pre-miRNA sequences processed by Dicer to yield mature miRNAs (Fig. 1D).

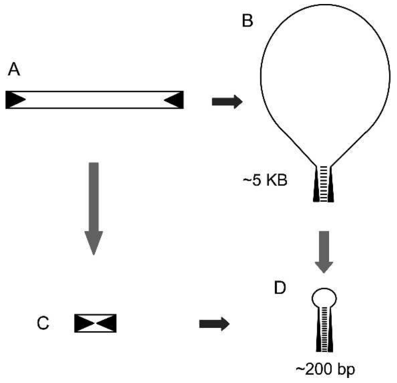

FIGURE 1. Model for the TE-based siRNA-miRNA evolutionary transition. (*A*) Full-length DNA-type element with terminal inverted repeats (TIRs) flanking a long open reading frame. (*B*) Snap-back secondary structure of the full-length element with TIRs bound as dsRNA. (*C*) MITE, a non-autonomous derivative of a full-length DNA-type element, containing TIRs and a small internal region. (*D*) Hairpin (stem–loop) secondary structure formed by a MITE RNA.

The relationship between full-length DNA-type elements and siRNAs, on the one hand, and MITEs and miRNAs, on the other, led us to propose a specific model for how miRNAs could have evolved from siRNA encoding TEs in a step-wise manner (Piriyapongsa and Jordan 2007). As illustrated in Figure 1, our model posits that siRNAs were first processed from the two TIRs of full-length elements bound as dsRNA. Later, as derivative MITEs evolved from full-length elements and proliferated in the genome, the same RNA endonucleolytic processing machinery cleaved the dsRNA from the hairpin stem regions, yielding mature miRNA sequences. A corollary prediction of our model holds that evolutionary intermediates may exist as TE sequences that encode both siRNAs and miRNAs. We tested the prediction of dual coding siRNA–miRNA TEs using a computational analysis of genome sequences, annotation and expression data from the plants *Arabidopsis thaliana* and *Oryza sativa* (rice).

RESULTS

We searched *Arabidopsis* and rice genome sequence and expression data (see Materials and Methods) to determine whether there are individual TE insertions that encode both siRNA and miRNA sequences. The *Arabidopsis* and rice genome sequences, along with their functional genomic data sets, afford several specific advantages for this kind of search. Both model species have been studied extensively, particularly by biologists interested in TEs, and accordingly their TEs are relatively well characterized. In addition, RNA expression levels for *Arabidopsis* and rice genes have been extensively characterized using the high-throughput massively parallel signature sequencing (MPSS) technique (Brenner et al. 2000a,b). The original MPSS technique was later modified to characterize small RNA sequences such as siRNAs and miRNAs (Lu et al. 2006). MPSS for short RNAs yields many thousands of sequence tags that can be unambiguously mapped to the *Arabidopsis* or rice genomes to determine where mature siRNAs and miRNAs are encoded.

The miRBase Sequence Database (Griffiths-Jones et al. 2006) contains genome annotations for experimentally characterized miRNA gene sequences from a number of species, including *Arabidopsis* and rice. Release 10.0 of miRBase contains 184 *Arabidopsis* and 231 rice miRNAs. We compared the genomic locations of these miRNAs to the locations of TEs annotated using the RepeatMasker program. Twelve *Arabidopsis* miRNAs (6.5%) and 83 rice miRNAs (35.9%) were found to be colocated with TE sequences (Supplemental Table 1). Ten out of 12 TE colocated *Arabidopsis* miRNA sequences and 38 out of 83 TE colocated rice miRNA sequences share 100% of their sequences with TEs (Supplemental Fig. 1). The TE sequences were all annotated based on RepeatMasker scores well above the threshold for false positives (average SW score = 20,357). In other words, these data represent unequivocal cases of plant miRNA genes that have been derived from TE sequences (Table 1). These miRNAs are derived from members of a variety of TE sequence families, including gypsy- and copia-like LTR retroelements, but the vast majority are encoded by the short nonautonomous DNA-type TEs known as MITEs. MITE-derived miRNAs are particularly enriched in rice consistent with the genomic abundance of MITEs in this species (Jiang et al. 2004).

miRBase was used to count the number of orthologs for each *Arabidopsis* and rice miRNA. TE-derived miRNA genes in *Arabidopsis* and rice have fewer orthologs on average (0.07), i.e., they are less evolutionarily conserved, than nonrepetitive miRNAs (3.0), and the difference is highly significant (Student's *t*-test; $t = 18.8$, df = 13, $P = 2e\text{-}57$). This is similar to what is seen for many mammalian TE-derived miRNAs (Piriyapongsa et al. 2007) and is consistent with the fact that TEs represent the most lineage-specific and rapidly evolving sequences in eukaryotic genomes (Marino-Ramirez et al. 2005). On the one hand, this may suggest that caution is warranted when evaluating TE-derived plant miRNAs since they are not conserved (Ambros et al. 2003). However, there are a number of bona fide miRNAs that are not evolutionarily conserved (Bentwich et al. 2005). The low conservation of TE-derived miRNAs can be taken to imply that the regulatory effects exerted by TE-derived miRNAs may be relevant for species-specific differences in gene expression (Piriyapongsa et al. 2007).

In addition to using miRBase to characterize TE-derived miRNAs, we searched the literature to confirm TE-derived plant miRNA genes with documented effects on the expression of host genes. There are five TE-derived miRNAs uncovered here (Table 1, ath-MIR854a–ath-MIR854d, ath-MIR855), including a repetitive family derived from dispersed LTR sequences, with experimentally characterized effects on the regulation of *Arabidopsis* genes (Arteaga-Vazquez et al. 2006). First of all, mature ath-MIR854 sequences were found to be absent in plants with mutant alleles for three genes critical to miRNA biogenesis: *Dicer-like1* (*dcl1*), *Hyponastic leaves1* (*hyl1*), and *HUA Enhancer1* (*hen1*). However, ath-MIR854 expression was found in mutants of the *RNA-dependent RNA polymerase2* gene, which is required for siRNA processing. Together, these results indicate that ath-MIR854 is processed specifically as an miRNA. The mature sequences of ath-MIR854 and ath-MIR855 have multiple binding sites in the 3′-untranslated region (UTR) of the *oligouridylate binding protein1b* gene (*UBP1b*), which encodes a heterogeneous nuclear RNA binding protein. The *UBP1b* 3′-UTR mRNA–miRNA interactions resemble those that lead to translational repression and/or mRNA cleavage in mammals. The ability of these TE-derived miRNAs to repress expression of *UBP1b* was demonstrated by using the 3′-UTR of the gene in a reporter protein expression assay. Mature ath-MIR854 and ath-MIR855 sequences are expressed in rosette leaves and flowers but absent in cauline leaves. Accordingly, the 3′-UTR of *UBP1b* can repress protein reporter expression in rosette leaves and flowers not in cauline leaves. Furthermore, comparison of mRNA versus protein expression for the reporter indicated that the ath-MIR854a–ath-MIR854d and ath-MIR855 genes exert their effects at the translational level.

We used expression data taken from the *Arabidopsis* MPSS Plus (Meyers et al. 2004) and Rice MPSS Plus (Nobuta et al. 2007) databases to evaluate whether any of the TEs that encode miRNA genes are also processed to yield siRNA sequences. The siRNA MPSS sequence tags were unambiguously mapped, using 100% tag-TE sequence identity, to the TEs that were found to encode miRNAs. Eight of the *Arabidopsis* TEs that encode miRNAs and 13 of the miRNA encoding rice TEs were found to encode siRNA sequences as well (Table 1).

We also explored the possibility that TE-derived miRNAs and siRNAs are passively transcribed as part of longer host gene RNA messages. To do this, genome-wide EST and mRNA maps for *Arabidopsis* and rice were examined to look for cases where dual coding miRNA–siRNA TE sequences are located within the start and stop coordinates of spliced transcripts. We were able to find several examples of such cases, one for *Arabidopsis* and two for rice (Fig. 2).

Finally, all of the miRNA–siRNA dual coding TE sequences were folded to predict their secondary structures (Supplemental Fig. 2). The predicted secondary structures show long double-stranded regions that correspond to the locations of mapped siRNA sequence tags along with stem–loop regions characteristic of known miRNA gene structures. The MITE encoded secondary structures are particularly striking in the sense that they form long, almost

TABLE 1. Plant miRNA genes derived from TEs

Name[a]	Accn[b]	Coords[c]	TE[d]	TE size[e]
ath-MIR855	MI0005411	chr2:4681509–4681780(+)	Athila4B_LTR (LTR/Gypsy)	Fragment
ath-MIR416	MI0001427	chr2:7015602–7015681(+)	Vandal1 (DNA/MuDR)	Fragment
ath-MIR405a	MI0001074	chr2:9642037–9642193(−)	SIMPLEHAT2 (DNA/hAT)	Fragment
ath-MIR405d	MI0001077	chr4:2789653–2789738(−)	SIMPLEHAT2 (DNA/hAT)	Fragment
ath-MIR401	MI0001070	chr4:5020234–5020483(−)	Athila4B_LTR (LTR/Gypsy)	Fragment
ath-MIR854b	MI0005413	chr5:11341600–11341820(−)	Athila6A_I (LTR/Gypsy)	Intact
ath-MIR854d	MI0005415	chr5:11707091–11707311(−)	Athila6A_I (LTR/Gypsy)	Intact
ath-MIR854c	MI0005414	chr5:11855326–11855546(+)	Athila6A_I (LTR/Gypsy)	Intact
ath-MIR854a	MI0005412	chr5:11864949–11865169(+)	Athila6A_I (LTR/Gypsy)	Intact
ath-MIR405b	MI0001075	chr5:20649740–20649863 (+)	SIMPLEHAT2 (DNA/hAT)	Fragment
osa-MIR439a	MI0001691	chr1:20206990–20207082(+)	MuDR4_OS (DNA/MuDR)	Fragment
osa-MIR814a	MI0005239	chr1:22701877–22701973(+)	STOWAWAY47_OS (DNA/Stowaway)	Intact
osa-MIR812a	MI0005233	chr1:34273999–34274232(+)	STOWAWAY51_OS (DNA/Stowaway)	Intact
osa-MIR819a	MI0005252	chr1:41534243–41534367(+)	STOWAWAY1_OS (DNA/Stowaway)	Intact
osa-MIR812b	MI0005234	chr2:1936324–1936493(−)	STOWAWAY51_OS (DNA/Stowaway)	Intact
osa-MIR818b	MI0005248	chr2:4007187–4007299(+)	STOWAWAY15–2_OS (DNA/Stowaway)	Intact
osa-MIR806b	MI0005211	chr2:5044109–5044323(−)	TREP215 (DNA/Stowaway)	Intact
osa-MIR814c	MI0005241	chr2:10889670–10889752(−)	STOWAWAY47_OS (DNA/Stowaway)	Fragment
osa-MIR817	MI0005246	chr2:12276361–12276443(−)	ENSPM3_OS (DNA/En-Spm)	Fragment
osa-MIR807b	MI0005218	chr2:24481931–24482076(−)	ECR (DNA/Tourist)	Intact
osa-MIR814b	MI0005240	chr2:26335342–26335415(+)	STOWAWAY47_OS (DNA/Stowaway)	Intact
osa-MIR819d	MI0005255	chr3:10848548–10848699(−)	STOWAWAY1_OS (DNA/Stowaway), STOWAWAY10_OS (DNA/Stowaway)	Intact
osa-MIR821a	MI0005266	chr3:22928833–22929106(+)	ENSPM3_OS (DNA/En-Spm), OSTE22 (DNA)	Fragment
osa-MIR443	MI0001708	chr3:29972009–29972156(+)	STOWAWAY47_OS (DNA/Stowaway)	Intact
osa-MIR420	MI0001440	chr4:6098543–6098697(+)	TRUNCATOR2_OS (LTR/Gypsy)	Intact
osa-MIR416	MI0001436	chr4:17268776–17268884(+)	CPSC3_LTR (LTR/Copia)	Intact
osa-MIR807c	MI0005219	chr4:23886344–23886527(+)	ECR (DNA/Tourist)	Intact
osa-MIR442	MI0001707	chr4:32149607–32149839(+)	OLO24B (DNA/Tourist)	Almost full length
osa-MIR819f	MI0005257	chr4:35070636–35070779(−)	STOWAWAY50_OS (DNA/Stowaway)	Intact
osa-MIR819g	MI0005258	chr5:28003948–28004094(+)	STOWAWAY1_OS:DNA/Stowaway	Intact
osa-MIR819h	MI0005259	chr6:10052973–10053127(−)	STOWAWAY50_OS (DNA/Stowaway), SZ-66LTR (LTR/Gypsy)	Intact
osa-MIR811a	MI0005230	chr6:13901553–13901742(+)	TAMI2 (DNA)	Intact
osa-MIR812c	MI0005235	chr6:26259310–26259473(+)	STOWAWAY9_OS (DNA/Stowaway)	Intact
osa-MIR821b	MI0005267	chr7:16415531–16415817(+)	OSTE22 (DNA), TNR3_OS (DNA/En-Spm)	Fragment
osa-MIR812d	MI0005236	chr7:22393529–22393681(+)	STOWAWAY44_OS (DNA/Stowaway)	Intact
osa-MIR445a	MI0001709	chr7:28117531–28117798(+)	NDNA2TNA_OS (DNA/Tourist)	Intact
osa-MIR818e	MI0005251	chr7:28152738–28152962(−)	STOWAWAY21_OS (DNA/Stowaway)	Intact
osa-MIR531	MI0003204	chr8:1214013–1214093(−)	SC-1_int-int (LTR/Copia)	Fragment
osa-MIR812e	MI0005237	chr8:16268303–16268472(+)	STOWAWAY44_OS (DNA/Stowaway)	Intact
osa-MIR821c	MI0005268	chr8:19792287–19792552(−)	ENSPM3_OS (DNA/En-Spm), OSTE22 (DNA)	Almost full length
osa-MIR811b	MI0005231	chr10:2372014–2372203(+)	TAMI2 (DNA)	Intact
osa-MIR439b	MI0001692	chr10:5338996–5339055(+)	MuDR4_OS (DNA/MuDR)	Fragment
osa-MIR816	MI0005245	chr10:21478646–21478722(+)	STOWAWAY47_OS (DNA/Stowaway)	Almost full length
osa-MIR806g	MI0005216	chr10:22588399–22588638(+)	TREP215 (DNA/Stowaway)	Intact
osa-MIR811c	MI0005232	chr11:5200383–5200541(−)	TAMI2 (DNA)	Fragment
osa-MIR813	MI0005238	chr11:23113437–23113639(+)	NDNA1TNA_OS (DNA/Tourist)	Fragment
osa-MIR531	MI0003204	chr11:26423868–26423948(+)	SC-1_int-int (LTR/Copia)	Intact
osa-MIR809h	MI0005228	chr12:5776955–5777088(+)	STOWAWAY1_OS (DNA/Stowaway)	Intact

[a] MiRBase database miRNA names.
[b] MiRBase database miRNA accessions.
[c] Genomic location coordinates of colocated TE sequences.
[d] Name, class, and family of the TE sequences.
[e] Size of TE sequences that encode miRNA genes: intact (full-length element), almost full-length (≥80% of full-length element consensus sequence), and fragment (<80% of full-length element consensus sequence).

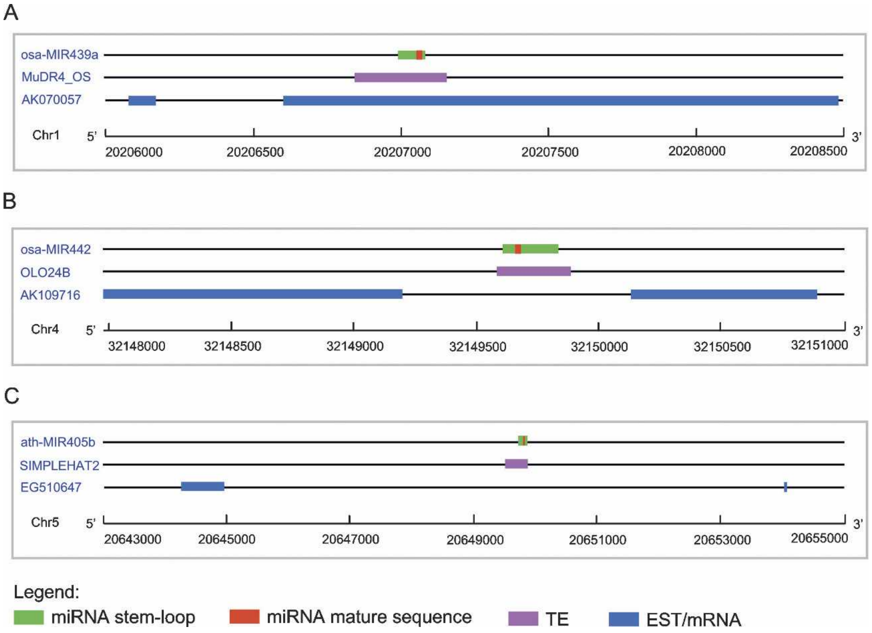

FIGURE 2. Genomic structure and expression of TE-derived miRNAs. The schematic diagrams representing the colocation between miRNA genes, TEs, and EST/mRNAs (see legend) are shown for (*A*) osa-MIR439a, (*B*) osa-MIR422, and (*C*) ath-MIR405b.

perfect hairpins possessing extensive double-stranded regions (Fig. 3; Supplemental Fig. 2). These folding patterns are based on the sequence complementarity between the TIRs encoded by this class of TEs.

DISCUSSION

Our analysis of *Arabidopsis* and rice genomic data revealed the existence of TE sequences that encode both siRNAs and miRNAs. We believe that the dual coding capacity for small regulatory RNAs by plant TEs reflects an evolutionary connection between related mechanisms of RNAi. This notion is based on the recent discovery of a family of human miRNA genes derived from MITEs, which led us to propose a step-wise model for the evolution of miRNAs from TEs that originally encoded siRNAs (Piriyapongsa and Jordan 2007). The expression of siRNAs from autonomous DNA-type elements is known to be based on read-through transcription of full-length elements (Sijen and Plasterk 2003), and our model is based on read-through transcription of shorter nonautonomous MITEs (Fig. 1). MITEs retain the TIR sequences of autonomous DNA-type elements but do not encode any open reading frame between the TIRs. As such, MITEs are made up mostly of TIR sequences; i.e., they are palindromic, and when expressed as read-through transcripts, they will fold to form hairpin structures similar to those of miRNA genes (Bartel 2004). Apparently, these MITE-derived hairpins can be processed to yield functionally relevant mature miRNA sequences (Piriyapongsa and Jordan 2007; Piriyapongsa et al. 2007). Consistent with this model, our results demonstrate that several of the siRNA–miRNA encoding TEs found in plants are in fact expressed as read-through transcripts by virtue of their presence in the introns of spliced RNA messages (Fig. 2, ESTs/mRNAs). After TE-containing introns are spliced from the mRNAs, they can fold to form the kinds of structures recognized by the endonucleases involved in RNAi (Fig. 3; Supplemental Fig. 2).

In addition to their palindromic sequence-structure characteristics, MITEs are also distinguished by their preference for insertion into gene rich regions (Mao et al. 2000; Zhang et al. 2000). Taken together with their genomic abundance (Jiang et al. 2004), this means that thousands of MITEs will be expressed as read-through transcripts as required by our model. The particular enrichment of MITEs in plant gene regions has been taken to suggest that they play some functional role for their host genomes. Our results, and our model of miRNA evolution via the autonomous TE-to-MITE transition, suggest that the host relevant function of MITEs is related, at least in part, to RNA-mediated gene regulation.

There is recent evidence from *Drosophila* in support of the notion that miRNAs can be processed from the introns of expressed genes in a manner similar to that which we propose for TE-derived miRNA genes (Ruby et al. 2007). Spliced introns that can fold into pre-miRNA-like structures and be cleaved to yield mature miRNA sequences are called "mirtrons." Interestingly, the processing of mirtrons to yield mature miRNAs does not rely on the RNA endonuclease Drosha. Drosha, or its plant functional analog Dicer-like1, is the enzyme that cleaves the longer pri-miRNA sequence near the base of the stem region in the nucleus to yield the pre-miRNA hairpin, which is then cleaved by Dicer to liberate the mature miRNA. Similar to the processing of mirtrons, PTGS via siRNAs does not require Drosha since dsRNAs are processed by Dicer alone to yield siRNAs (Bartel 2004). Thus, the work of Ruby et al. (2007) indicates that miRNAs can arise in any organism that possesses both spliceosomal introns and PTGS via siRNAs; this was taken to suggest that miRNAs may have emerged in ancient eukaryotes prior to the evolution of the complete miRNA biogenesis pathway. Our model points to MITEs as a potential source for the evolution of such ancient miRNAs, processed by Dicer (or Dicer-like1) alone, from full-length TEs that previously encoded siRNAs only. Consistent with an ancient origin of miRNAs from TEs, full-length DNA-type elements and MITEs are widely

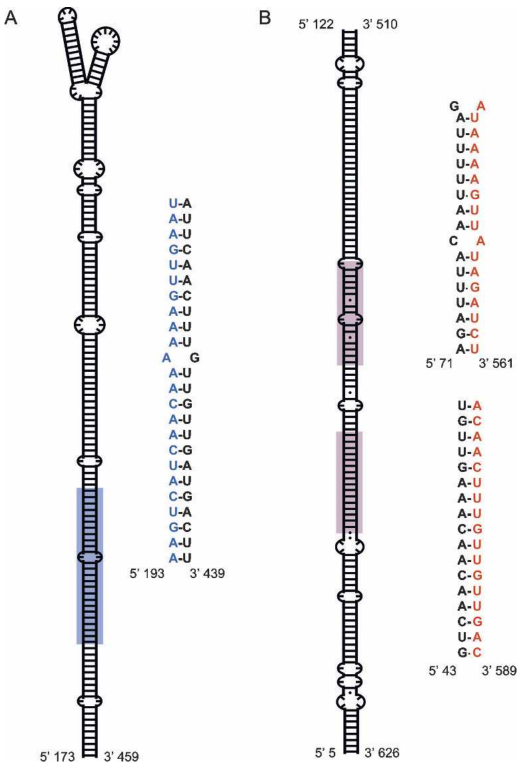

FIGURE 3. RNA secondary structure and sequences of an siRNA–miRNA dual encoding MITE sequence. Partial predicted secondary structures of a read-through transcript of a MITE encoding both siRNA and miRNA sequences are shown for the rice miRNA gene osa-MIR821b. (*A*) A schematic of the miRNA stem–loop region along with the miRNA mature sequence. The location of the miRNA mature sequence in the stem–loop is indicated with blue shading, and the mature miRNA sequence residues are shown in blue. (*B*) A schematic of the double-stranded RNA region that is cleaved to yield siRNAs. The locations of the siRNA signatures in the sequence are indicated with red shading, and the siRNA signature sequences are shown in red. Note that the entire secondary structure for this MITE is shown in Supplemental Figure 2R.

distributed among eukaryotes (Feschotte et al. 2002), indicating that they were likely to be present in ancestral eukaryotic species. In addition, a recent phylogenetic analysis of the miRNA biogenesis enzymes indicates that Dicer is the more ancient of the endonucleases involved the processing of mature miRNAs, with Drosha having evolved more recently along the animal evolutionary lineage (Cerutti and Casas-Mollano 2006).

dsRNA sequences are processed to yield multiple siRNAs from a given stretch of sequence, while pre-miRNA hairpins are cleaved into a single distinct mature miRNA sequence (Bartel 2004). This may be related to the steric hindrance entailed by the substantially shorter hairpin structures that are processed to yield miRNAs. Over evolutionary time, once the endonucleolytic machinery became tuned to the structural characteristics, and limited spacing, of the MITE-encoded hairpins, then it would have been able to recognize any number of hairpin structures that are formed when genomic sequences with inverted repeats are expressed as read-through transcripts. Indeed, this has been shown to be important in *Arabidopsis*, where miRNA genes evolved via local inverted duplication events, which generated sequences capable of folding back into hairpin structures when expressed (Allen et al. 2004). In this way, MITEs could have stimulated the RNAi biogenesis enzymes to process non-TE-related hairpin structures to yield miRNAs with host gene regulatory functions.

Relatively ancient siRNA sequences originally evolved as defense mechanisms against genomic invaders, such as viruses and TEs, and genome defense appears to remain the primary function of this class of regulatory sequence. On the other hand, miRNAs are evolutionarily emergent regulators, and accordingly they function primarily to regulate host genes. The siRNA-to-miRNA evolutionary transition is one of a growing number of examples (Yoder et al. 1997; Matzke et al. 2000; Lippman et al. 2004; McDonald et al. 2005) of gene silencing mechanisms that were originally employed to defend against TE proliferation and were later co-opted to serve the regulatory needs of the host organism (Piriyapongsa et al. 2007).

MATERIALS AND METHODS

The *A. thaliana* genomic sequence was obtained from the National Center for Biotechnology (NCBI) genome assembly/annotation projects ftp site (ftp://ftp.ncbi.nih.gov/genomes/Arabidopsis_thaliana). The *O. sativa* (rice) genomic sequence was taken from release 4.0 of The Institute for Genomic Research (TIGR) rice genome annotation database (Ouyang et al. 2007; ftp://ftp.tigr.org/pub/data/Eukaryotic_Projects/o_sativa/annotation_dbs/pseudo molecules/). The genome locations of different classes of TEs in *Arabidopsis* and rice genomes were identified by using the RepeatMasker program (Smit et al. 2004) to compare genomic sequences against the species-specific Repbase libraries (Jurka 2000; Jurka et al. 2005) of TE consensus sequences. The Smith–Waterman (SW) algorithm (Smith and Waterman 1981) was used with RepeatMasker to do local pairwise comparisons of genome-against-TE consensus sequences and to score the resulting alignments. The genome locations and identities of experimentally characterized *Arabidopsis* and rice miRNA gene sequences were taken from release 10.0 of the miRBase database (Griffiths-Jones et al. 2006; http://microrna.sanger.ac.uk/sequences/).

Arabidopsis and rice MPSS small RNA signatures were downloaded from the *Arabidopsis* MPSS Plus database (Meyers et al. 2004; http://mpss.udel.edu/at/) and the Rice MPSS Plus database (Nobuta et al. 2007; http://mpss.udel.edu/rice/). The signatures matching to tRNAs, rRNAs, snRNAs, or snoRNAs were not included in the data set we used. Only the small RNA signatures of size 17–25 bp were chosen for the analysis. For each species, the small RNA signatures were divided into two groups: miRNA signatures and siRNA signatures according the organism-specific MPSS database annotations.

The genome locations of TEs and miRNAs were compared to identify the co-located TEs and miRNA gene sequences in both species. Post-processing of RepeatMasker annotations was done such that the continuous TE sequences of the same family, which are oriented in the same direction on the genome, were counted as the same TE sequence. TE sequences that encoded entire miRNA gene sequences were searched for the presence of small RNA signatures using the vmatch program (Abouelhoda et al. 2004) demanding 100% sequence identity between the TE sequences and siRNA tags. The TE sequences that completely covered miRNA gene sequences and contained siRNA signatures outside the miRNA gene regions were chosen for further analysis. These TE sequences were folded using the program RNAfold from the Vienna RNA package (Hofacker et al. 1994), and their secondary structures were visualized by xrna program (http://rna.ucsc.edu/rnacenter/xrna/xrna.html). The potential of TE-derived miRNAs and siRNAs to be processed from read-through transcripts was assessed via the analysis of EST and mRNA data. EST and mRNA sequences mapped to the *Arabidopsis* genome sequence were obtained from NCBI genome assembly/annotation projects (ftp://ftp.ncbi.nih.gov/genomes/Arabidopsis_thaliana/GNOMON). Mapped rice EST, full-length cDNA sequences, and transcript assemblies were obtained from TIGR rice genome annotation database (Ouyang et al. 2007).

SUPPLEMENTAL DATA

Supplemental material can be found at http://www.rnajournal.org.

Received November 10, 2007; accepted February 15, 2008.

REFERENCES

Abouelhoda, M.I., Kurtz, S., and Ohlebusch, E. 2004. Replacing suffix trees with enhanced suffix arrays. *J. Discrete Algorithm* **2:** 53–86.

Allen, E., Xie, Z., Gustafson, A.M., Sung, G.H., Spatafora, J.W., and Carrington, J.C. 2004. Evolution of microRNA genes by inverted duplication of target gene sequences in *Arabidopsis thaliana*. *Nat. Genet.* **36:** 1282–1290.

Ambros, V. 2004. The functions of animal microRNAs. *Nature* **431:** 350–355.

Ambros, V., Bartel, B., Bartel, D.P., Burge, C.B., Carrington, J.C., Chen, X., Dreyfuss, G., Eddy, S.R., Griffiths-Jones, S., Marshall, M., et al. 2003. A uniform system for microRNA annotation. *RNA* **9:** 277–279.

Arteaga-Vazquez, M., Caballero-Perez, J., and Vielle-Calzada, J.P. 2006. A family of microRNAs present in plants and animals. *Plant Cell* **18:** 3355–3369.

Bartel, D.P. 2004. MicroRNAs: Genomics, biogenesis, mechanism, and function. *Cell* **116:** 281–297.

Bentwich, I., Avniel, A., Karov, Y., Aharonov, R., Gilad, S., Barad, O., Barzilai, A., Einat, P., Einav, U., Meiri, E., et al. 2005. Identification of hundreds of conserved and nonconserved human microRNAs. *Nat. Genet.* **37:** 766–770.

Borchert, G.M., Lanier, W., and Davidson, B.L. 2006. RNA polymerase III transcribes human microRNAs. *Nat. Struct. Mol. Biol.* **13:** 1097–1101.

Brenner, S., Johnson, M., Bridgham, J., Golda, G., Lloyd, D.H., Johnson, D., Luo, S., McCurdy, S., Foy, M., Ewan, M., et al. 2000a. Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. *Nat. Biotechnol.* **18:** 630–634.

Brenner, S., Williams, S.R., Vermaas, E.H., Storck, T., Moon, K., McCollum, C., Mao, J.I., Luo, S., Kirchner, J.J., Eletr, S., et al. 2000b. In vitro cloning of complex mixtures of DNA on microbeads: Physical separation of differentially expressed cDNAs. *Proc. Natl. Acad. Sci.* **97:** 1665–1670.

Bureau, T.E. and Wessler, S.R. 1992. Tourist: A large family of small inverted repeat elements frequently associated with maize genes. *Plant Cell* **4:** 1283–1294.

Bureau, T.E. and Wessler, S.R. 1994. Stowaway: A new family of inverted repeat elements associated with the genes of both monocotyledonous and dicotyledonous plants. *Plant Cell* **6:** 907–916.

Cerutti, H. and Casas-Mollano, J.A. 2006. On the origin and functions of RNA-mediated silencing: From protists to man. *Curr. Genet.* **50:** 81–99.

Covey, S.N., Al-Kaff, N.S., Lángara, A., and Turner, D.S. 1997. Plants combat infection by gene silencing. *Nature* **385:** 781–782.

de Carvalho, F., Gheysen, G., Kushnir, S., Van Montagu, M., Inze, D., and Castresana, C. 1992. Suppression of β-1,3-glucanase transgene expression in homozygous plants. *EMBO J.* **11:** 2595–2602.

Feschotte, C. and Mouches, C. 2000. 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.* **17:** 730–737.

Feschotte, C., Zhang, X., and Wessler, S.R 2002. Miniature inverted-repeat transposable elements (MITES) and their relationships to established DNA transposons. In Mobile DNA II (eds. N. Craig et al.), pp. 1147–1158. American Society for Microbiology Press, Washington, DC.

Fire, A., Xu, S., Montgomery, M.K., Kostas, S.A., Driver, S.E., and Mello, C.C. 1998. Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* **391:** 806–811.

Griffiths-Jones, S., Grocock, R.J., van Dongen, S., Bateman, A., and Enright, A.J. 2006. miRBase: MicroRNA sequences, targets, and gene nomenclature. *Nucleic Acids Res.* **34:** D140–D144. doi: 10.1093/nar/gkj112.

Hofacker, I., Fontana, W., Stadler, P., Bonhoeffer, S., Tacker, M., and Schuster, P. 1994. Fast folding and comparison of RNA secondary structures. *Monatsh. Chem.* **125:** 167–188.

Jiang, N., Feschotte, C., Zhang, X., and Wessler, S.R. 2004. Using rice to understand the origin and amplification of miniature inverted repeat transposable elements (MITEs). *Curr. Opin. Plant Biol.* **7:** 115–119.

Jurka, J. 2000. Repbase update: A database and an electronic journal of repetitive elements. *Trends Genet.* **16:** 418–420.

Jurka, J., Kapitonov, V.V., Pavlicek, A., Klonowski, P., Kohany, O., and Walichiewicz, J. 2005. Repbase Update, a database of eukaryotic repetitive elements. *Cytogenet. Genome Res.* **110:** 462–467.

Ketting, R.F., Haverkamp, T.H., van Luenen, H.G., and Plasterk, R.H. 1999. Mut-7 of *C. elegans*, required for transposon silencing and RNA interference, is a homolog of Werner syndrome helicase and RNaseD. *Cell* **99:** 133–141.

Lindbo, J.A., Silva-Rosales, L., Proebsting, W.M., and Dougherty, W.G. 1993. Induction of a highly specific antiviral state in transgenic plants: Implications for regulation of gene expression and virus resistance. *Plant Cell* **5:** 1749–1759.

Lippman, Z., Gendrel, A.V., Black, M., Vaughn, M.W., Dedhia, N., McCombie, W.R., Lavine, K., Mittal, V., May, B., Kasschau, K.D., et al. 2004. Role of transposable elements in heterochromatin and epigenetic control. *Nature* **430:** 471–476.

Lu, C., Kulkarni, K., Souret, F.F., MuthuValliappan, R., Tej, S.S., Poethig, R.S., Henderson, I.R., Jacobsen, S.E., Wang, W., Green, P.J., et al. 2006. MicroRNAs and other small RNAs enriched in the *Arabidopsis* RNA-dependent RNA polymerase-2 mutant. *Genome Res.* **16:** 1276–1288.

Mao, L., Wood, T.C., Yu, Y., Budiman, M.A., Tomkins, J., Woo, S., Sasinowski, M., Presting, G., Frisch, D., Goff, S., et al. 2000. Rice transposable elements: A survey of 73,000 sequence-tagged-connectors. *Genome Res.* **10:** 982–990.

Marino-Ramirez, L., Lewis, K.C., Landsman, D., and Jordan, I.K. 2005. Transposable elements donate lineage-specific regulatory sequences to host genomes. *Cytogenet. Genome Res.* **110:** 333–341.

Matzke, M.A. and Matzke, A.J. 2004. Planting the seeds of a new paradigm. *PLoS Biol.* **2:** E133. doi: 10.1371/journal.pbio.0020133.

Matzke, M.A., Mette, M.F., and Matzke, A.J. 2000. Transgene silencing by the host genome defense: Implications for the evolution of epigenetic control mechanisms in plants and vertebrates. *Plant Mol. Biol.* **43:** 401–415.

McDonald, J.F., Matzke, M.A., and Matzke, A.J. 2005. Host defenses to transposable elements and the evolution of genomic imprinting. *Cytogenet. Genome Res.* **110:** 242–249.

Mette, M.F., van der Winden, J., Matzke, M., and Matzke, A.J. 2002. Short RNAs can identify new candidate transposable element families in *Arabidopsis*. *Plant Physiol.* **130:** 6–9.

Meyers, B.C., Lee, D.K., Vu, T.H., Tej, S.S., Edberg, S.B., Matvienko, M., and Tindell, L.D. 2004. Arabidopsis MPSS. An online resource for quantitative expression analysis. *Plant Physiol.* **135:** 801–813.

Morgan, G.T. 1995. Identification in the human genome of mobile elements spread by DNA-mediated transposition. *J. Mol. Biol.* **254:** 1–5.

Napoli, C., Lemieux, C., and Jorgensen, R. 1990. Introduction of a chimeric chalcone synthase gene into petunia results in reversible co-suppression of homologous genes in trans. *Plant Cell* **2:** 279–289.

Nobuta, K., Venu, R.C., Lu, C., Belo, A., Vemaraju, K., Kulkarni, K., Wang, W., Pillay, M., Green, P.J., Wang, G.L., et al. 2007. An expression atlas of rice mRNAs and small RNAs. *Nat. Biotechnol.* **25:** 473–477.

Oosumi, T., Belknap, W.R., and Garlick, B. 1995. Mariner transposons in humans. *Nature* **378:** 672.

Ouyang, S., Zhu, W., Hamilton, J., Lin, H., Campbell, M., Childs, K., Thibaud-Nissen, F., Malek, R.L., Lee, Y., Zheng, L., et al. 2007. The TIGR Rice Genome Annotation Resource: Improvements and new features. *Nucleic Acids Res.* **35:** D883–D887. doi: 10.1093/nar/gkl976.

Piriyapongsa, J. and Jordan, I.K. 2007. A family of human microRNA genes from miniature inverted-repeat transposable elements. *PLoS ONE* **2:** e203. doi: 10.1371/journal.pone.0000203.

Piriyapongsa, J., Marino-Ramirez, L., and Jordan, I.K. 2007. Origin and evolution of human microRNAs from transposable elements. *Genetics* **176:** 1323–1337.

Plasterk, R.H. 2002. RNA silencing: The genome's immune system. *Science* **296:** 1263–1265.

Ratcliff, F., Harrison, B.D., and Baulcombe, D.C. 1997. A similarity between viral defense and gene silencing in plants. *Science* **276:** 1558–1560.

Ruby, J.G., Jan, C.H., and Bartel, D.P. 2007. Intronic microRNA precursors that bypass Drosha processing. *Nature* **448:** 83–86.

Sijen, T. and Plasterk, R.H. 2003. Transposon silencing in the *Caenorhabditis elegans* germ line by natural RNAi. *Nature* **426:** 310–314.

Slotkin, R.K., Freeling, M., and Lisch, D. 2005. Heritable transposon silencing initiated by a naturally occurring transposon inverted duplication. *Nat. Genet.* **37:** 641–644.

Smalheiser, N.R. and Torvik, V.I. 2005. Mammalian microRNAs derived from genomic repeats. *Trends Genet.* **21:** 322–326.

Smit, A.F. and Riggs, A.D. 1996. Tiggers and DNA transposon fossils in the human genome. *Proc. Natl. Acad. Sci.* **93:** 1443–1448.

Smit, A.F., Hubley, R., and Green, P. 2004. RepeatMasker Open-3.0. http://www.repeatmasker.org/.

Smith, T.F. and Waterman, M.S. 1981. Identification of common molecular subsequences. *J. Mol. Biol.* **147:** 195–197.

Tabara, H., Sarkissian, M., Kelly, W.G., Fleenor, J., Grishok, A., Timmons, L., Fire, A., and Mello, C.C. 1999. The rde-1 gene, RNA interference, and transposon silencing in *C. elegans*. *Cell* **99:** 123–132.

van Blokland, R., van der Geest, N., Mol, J.N.M., and Kooter, J.M. 1994. Transgene-mediated suppression of chalcone synthase expression in Petunia hybrida results from an increase in RNA turnover. *Plant J.* **6:** 861–877.

van der Krol, A.R., Mur, L.A., Beld, M., Mol, J.N., and Stuitje, A.R. 1990. Flavonoid genes in petunia: Addition of a limited number of gene copies may lead to a suppression of gene expression. *Plant Cell* **2:** 291–299.

Yoder, J.A., Walsh, C.P., and Bestor, T.H. 1997. Cytosine methylation and the ecology of intragenomic parasites. *Trends Genet.* **13:** 335–340.

Zhang, Q., Arbuckle, J., and Wessler, S.R. 2000. 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.* **97:** 1160–1165.