Identification and characterization of endogenous small interfering RNAs from rice

Ramanjulu Sunkar, Thomas Girke and Jian-Kang Zhu*

Department of Botany and Plant Sciences, Institute for Integrative Genome Biology, 2150 Batchelor Hall, University of California, Riverside, CA 92521, USA

ABSTRACT

RNA silencing-mediated small interfering RNAs (siRNAs) and microRNAs (miRNAs) have diverse natural roles, ranging from regulation of gene expression and heterochromatin formation to genome defense against transposons and viruses. Unlike miRNAs, endogenous siRNAs are generally not conserved between species; consequently, their identification requires experimental approaches. Thus far, endogenous siRNAs have not been reported from rice, which is a model species for monocotyledonous plants. We identified a large set of putative endogenous siRNAs from root, shoot and inflorescence small RNA cDNA libraries of rice. Most of these siRNAs are from intergenic regions, although a substantial proportion (22%) originates from the introns and exons of protein-coding genes. Northern and RT–PCR analysis revealed that the expression of some of the siRNAs is tissue specific or developmental stage specific. A total of 25 transposons and 21 protein-coding genes were predicted to be *cis*-targets of some of the siRNAs. Based on sequence homology, we also predicted 111 putative *trans*-targets for 44 of the siRNAs. Interestingly, ∼46% of the predicted *trans*-targets are transposable elements, which suggests that endogenous siRNAs may play an important role in the suppression of transposon proliferation. Using RNA ligase-mediated-5′ rapid amplification of cDNA end assays, we validated three of the predicted targets and provided evidence for both *cis*- and *trans*-silencing of target genes by siRNAs-guided mRNA cleavage.

INTRODUCTION

In plants and animals, microRNAs (miRNAs) and small interfering RNAs (siRNAs) have emerged as important negative regulators of gene expression (1–4). miRNAs arise by endonucleolytic processing by the enzyme Dicer from hairpin-structured single-stranded precursor RNAs that are transcribed from endogenous nonprotein-coding genes (2,3). siRNAs are also processed by a Dicer, but are generated from double-stranded RNAs (dsRNAs) as a result of antisense or convergent transcription or due to the activity of one or more cellular RNA-dependent RNA polymerases (RdRPs) (4). These small RNAs serve as specificity determinants of transcriptional gene silencing (TGS) and post-transcriptional gene silencing (PTGS) (5–7).

The RNAi pathway and machinery are highly conserved throughout lower to higher eukaryotic organisms (*Schizosaccharomyces pombe* to metazoans). However, there are differences in siRNA-mediated RNA silencing pathways between plants and animals. For instance, siRNAs produced in *Drosophila* embryos (7) and in mammalian cells (8) belong to the ∼21 nt class, while siRNAs in plants and fungi fall into two distinct classes: a short (∼21 nt) and a long (∼24 nt) size class (9–15). Recently, it was shown that the ∼21 nt class of siRNAs designated as *trans*-acting siRNAs (tasiRNAs) from *Arabidopsis* are associated with post-transcriptional silencing by directing the cleavage of target mRNAs (16,17). The longer size class of siRNAs (∼24 nt) are associated with TGS involving DNA methylation and histone (H3K9) methylation (18–20). The length and functional diversity of small RNAs in plants are reflected in the multiplicity of DCL (Dicer-like) activities. Dicer is represented by one or two genes in animals, which indicates that often a single Dicer processes both miRNAs and siRNAs. In contrast, *Arabidopsis* and rice encodes at least four and three DCL proteins, respectively (19). *Arabidopsis* homozygous for the weak loss-of-function allele *dcl1-9* is impaired in miRNA precursor processing (15,21–23), whereas the DCL2 and DCL3 proteins are implicated in viral siRNA biogenesis and endogenous siRNA biogenesis, respectively (19). In addition to DCLs, an Argonaute (Ago-4), HEN1 and SDE4 have been implicated in siRNA accumulation (18,24–26). These studies suggest that plants have evolved multiple small

---

*To whom correspondence should be addressed. Tel: +1 909 827 7117; Fax +1 909 827 7115; Email: jian-kang.zhu@ucr.edu

RNA processing pathways with specific as well as overlapping functions.

Cloning of small RNAs is a starting point to understand their number, diversity and possible roles in any organism. Recent studies clearly indicated the importance of small RNA cloning, particularly in the identification of hitherto unknown classes of endogenous small RNAs in diverse species, such as *S.pombe*, *Drosophila*, *Caenorhabditis elegans* and *Arabidopsis*. A large number of endogenous small RNAs referred to as repeat-associated small interfering RNAs (rasiRNAs) that can be mapped to transposons of the *Drosophila* genome have been identified (27). Another class of small RNAs referred to as tiny non-coding RNAs (tncRNAs) have been identified in *C.elegans* (28). Heterochromatic siRNAs that corresponded to both DNA strands of centromeric repeats have been identified in *S.pombe* (29). Similarly, cloning in *Arabidopsis* led to the identification of a large number of endogenous siRNAs (15,16,19). The functions of these recently discovered endogenous siRNAs have not been investigated in detail, but they appear to have important regulatory roles in gene expression. For example, two recent studies have demonstrated that *Arabidopsis* endogenous siRNAs referred to as tasiRNAs act similar to miRNAs and can direct the cleavage of the predicted *trans*-target mRNAs (16,17).

Since computational methods to predict endogenous siRNAs are currently unavailable, the identification of these siRNAs requires experimental approaches. Endogenous siRNAs have thus far been cloned only from a dicotyledonous plant (*Arabidopsis*), but siRNAs from monocotyledonous plants are not known. Rice is the world's most important crop, as measured by the portion of calories provided to the human diet, and is the only monocot species with a completely sequenced genome. To identify endogenous siRNAs and miRNAs from rice, we generated three small RNA cDNA libraries from different tissues. In this study, we report the first identification and characterization of a large set of putative endogenous siRNAs from rice. Most of these siRNAs map to intergenic regions. Some of the siRNAs show tissue/developmental stage-dependent expression. A total of 5 of the siRNAs are perfectly complementary to 25 different transposons and 8 of the siRNAs display perfect complementarity with 21 protein-coding genes, suggesting the possibility that these genes might be regulated in *cis* by endogenous siRNAs. Besides these, 111 *trans*-targets were predicted for 44 of the siRNAs and the predicted *trans*-target genes consist largely of transposable elements. This observation implies a role for endogenous siRNAs in controlling the proliferation of transposons in both *cis*- and *trans*-silencing manners. Importantly, we show that three of the predicted targets are genuine targets of endogenous siRNAs in rice and their mRNA cleavage *in vivo* is guided by the siRNAs. This is the first demonstration of mRNA targets of endogenous siRNAs in rice.

MATERIALS AND METHODS

Cloning of endogenous siRNAs from rice

Total RNA was isolated separately from shoots and roots of 4-week-old young seedlings and inflorescences of adult rice plants (*Oryza sativa* spp. *japonica* cv. Nipponbare) with use of TRIzol (Invitrogen Life Technologies, Carlsbad, CA) according to the manufacturer's instructions. Cloning of small RNAs was performed as described previously (15). In brief, small RNAs from 18 to 26 nt were size fractionated, purified and ligated sequentially to 5′ and 3′ RNA/DNA chimeric oligonucleotide adapters. Reverse transcription was performed after ligation with the adapters, followed by PCR amplification. The resulting PCR products were cloned and transformed into competent cells. Plasmids were isolated from individual colonies and sequenced.

Sequence analysis

Automated base calling of raw sequence traces and vector removal were performed with the PHRED and CROSS MATCH programs from Ewing and Green (30). To avoid loss of critical sequence information in our short RNA sequences, a two step approach was used. First, non-insert and low quality sequences were removed with relatively forgiving filter settings by running the CROSS MATCH program with a minimum match parameter of 15 nt and a PHRED score of 14. In a second step, the obtained insert sequences with quality scores below 20 were flagged and further quality inspected by eye. Candidates with ambiguous base calls were removed from the dataset. The filtered and trimmed sequences with >16 nt in length were used to search the Rfam database (31) using the BLASTN program (32). This step allowed the removal of most RNA species that might be degradation products of non-coding RNAs in the dataset. Putative origins for the remaining sequences were identified by BLASTN searches against the genomic sequences of the version 2.0 annotation of *O.sativa* spp. *japonica* from The Institute for Genomic Research (TIGR) (ftp://ftp.tigr.org/pub/data/Eukaryotic_Projects/o_sativa/). Candidates with perfect matches against these genomic datasets were designated as endogenous siRNAs when they showed no fold-back structure in secondary structure predictions with the mfold program (33).

To identify putative target sequences, all predicted and cloned CDS (coding sequences) and UTR sequences from *O.sativa* (TIGR all.cdna set) were searched using the PatScan program (34). The following parameters were used for these pattern searches, all referring to the 5′ end of the siRNAs in antisense orientation: no more than three mismatches excluding at the positions 10 and 11. The retrieved siRNA/target site pairs were ranked and scored by aligning them with the Needleman–Wunsch global alignment program from EMBOSS (35).

RNA blot analysis

Total RNA was isolated from root, leaf and inflorescence tissues of adult plants and from 4-week-old young seedlings with use of Trizol. Low molecular weight RNAs were isolated by polyethylene glycol precipitation of total RNA (5). Fifty micrograms of low molecular weight RNA was loaded per lane, resolved on a denaturing 15% polyacrylamide gel and transferred electrophoretically to Hybond-N$^+$ membranes. Hybridization and washings were performed as described previously (15). The membranes were briefly air dried and then exposed to a PhosphorImager.

PCR-based detection of candidate siRNAs

PCR-based detection of small RNAs was performed as described previously by Grad *et al.* (36). To PCR amplify the candidate siRNAs from the small RNA libraries, an oligonucleotide complementary to the 5′ adapter was used with a 3′ oligonucleotide complementary to the particular candidate siRNA.

5′ Rapid amplification of cDNA ends

Total RNA from 4-week-old rice seedlings was extracted with use of Trizol reagent. Poly(A)$^+$ mRNA was purified from total RNA with use of the Poly(A) kit (Promega). RLM-5′ RACE (RNA ligase-mediated-5′ rapid amplification of cDNA ends) was carried out with use of the GeneRacer Kit (Invitrogen Life Technologies). The GeneRacer RNA Oligo adapter was directly ligated to mRNA (100 ng) without calf intestinal phosphatase and tobacco acid pyrophosphatase treatment. Gene-specific primers (9629.m00201, 9636.m02299 and 9632.m00807) were designed and used for cDNA synthesis. Initial PCR was carried out with the GeneRacer 5′ primer and gene specific primers. Nested PCR was carried out with 1 μl of the initial PCR, GeneRace 5′ nested primer and gene-specific internal primers. After amplification, 5′ RACE products were gel-purified and cloned, and at least 10 independent clones were sequenced.

RESULTS

Cloning, size distribution and 5′ nucleotide preference of rice endogenous siRNAs

We constructed three libraries of small RNAs each from the shoots and roots of seedlings and from the inflorescence tissues of adult rice plants (*O.sativa* spp. *japonica* cv. Nipponbare) plants. Small RNAs from 16 to 26 nt in length were isolated by size fractionation and ligated to 5′ and 3′ adapters, and then cloned and sequenced. A total of ∼10 000 clones were sequenced (∼1/3 from each library) and nearly half of it was represented by clones either shorter than 16 nt or without cDNA inserts (i.e. self-ligation of the adapters). The remaining 4910 small cDNA sequences were between 16 and 26 nt in length. All vector trimmed sequences longer than 16 nt were used to search the Rfam database (www.sanger.ac.uk/ Software/Rfam) with BLASTN (31). This step allowed for the removal of most RNA species that might be degradation products of non-coding RNAs in the dataset (Table 1). The largest class of cloned RNAs represents fragments of abundant non-coding RNAs (rRNA, tRNA, snRNA and snoRNA) as determined by BLASTN searches against the Rfam database. Seventy percent of the sequences correspond to ribosomal RNAs. Furthermore, 26S rRNAs and 18S rRNAs together accounted for ∼40% of the total sequences (Table 1). Several clones mapped to chloroplast or mitochondrial genomes and may represent either degradation or possibly regulatory products of organellar RNAs (Table 1). Several sequences (478) did not match to any nuclear or organellar sequence, suggesting that these may correspond to unfinished regions in the rice genome (Table 1). Putative origins for the remaining sequences were identified by BLASTN searches against the genomic sequences of *O.sativa* spp. *japonica*. Candidates with

Table 1. Small RNA composition of the cDNA library

RNA class	Numbers of small RNAs
28S rRNA	189
26S rRNA	1218
25S rRNA	407
23S rRNA	292
18S rRNA	846
16S rRNA	307
5S rRNA	291
tRNA	253
SnRNA	52
SnoRNA	18
Chloroplast/mitochondria	121
miRNAs	124
Endogenous siRNAs	314
No perfect hits to rice genome	478
Total	4910

perfect matches against these genomic datasets were endogenous small RNAs. These consist of miRNAs (124 sequences) (33) and putative endogenous siRNAs (314 sequences representing 284 unique siRNAs; all sequences are provided in Supplementary Table 1). The putative endogenous siRNAs correspond to genomic sequences that do not have predictable fold-back structures. A large set of small RNAs with sizes between 20 and 24 nt and in the sense orientation to non-coding RNAs (rRNA, tRNA, snRNA and snoRNA) and exons of protein-coding genes were not included in this analysis. Although many of these are likely degradation products of the abundant non-coding RNAs or mRNAs, some may be siRNAs. For example, some of the endogenous siRNAs in *Arabidopsis* have been shown to be derived from rDNAs and exons of protein-coding genes in the sense orientation (12,19). However, these were not characterized further in this study. Ascertaining these molecules unambiguously as rice siRNAs will require their expression analysis in dcl and rdrp mutants, which are not available for this species at the present time.

The number of endogenous siRNAs cloned from the different libraries varied greatly (Figure 1A). Most of them (225) are derived from the inflorescence library, whereas the shoot and root libraries yielded a substantially lower number of siRNAs (42 and 17 siRNAs from shoot and root libraries, respectively) (Figure 1A). Fifteen of the siRNA sequences (P4-D6, P8-E2, P16-C3, P65-B7, P86-H10, P88-A8, P88-A11, P91-A10, P94-H11, P98-A9, P103-B2, P104-G7, P107-E1, P108-A7 and P108-D3) showed an overlap between the libraries. The majority of the siRNAs are 21–24 nt in size, which is the typical size range for Dicer-derived products (Figure 1B). The 24 nt size class is predominant. The identified siRNAs exhibit a biased nucleotide distribution at their 5′ end. Of the 284 unique siRNAs, 105 begin with a U and 95 with an A, regardless of their size. About two-thirds begin with a pyrimidine, whereas the sequences that begin with C or G are underrepresented among the cloned sequences (Supplementary Table 1). Furthermore, each of the two size classes start with a distinct nucleotide at their 5′ end, with U predominating in the 21 nt class and a 5′ A in the 24 nt class (Figure 1D).

The publicly available *Arabidopsis* siRNAs (http://asrp. cgrb.oregonstate.edu/) were analyzed similarly to determine whether this nucleotide preference is a common feature in the
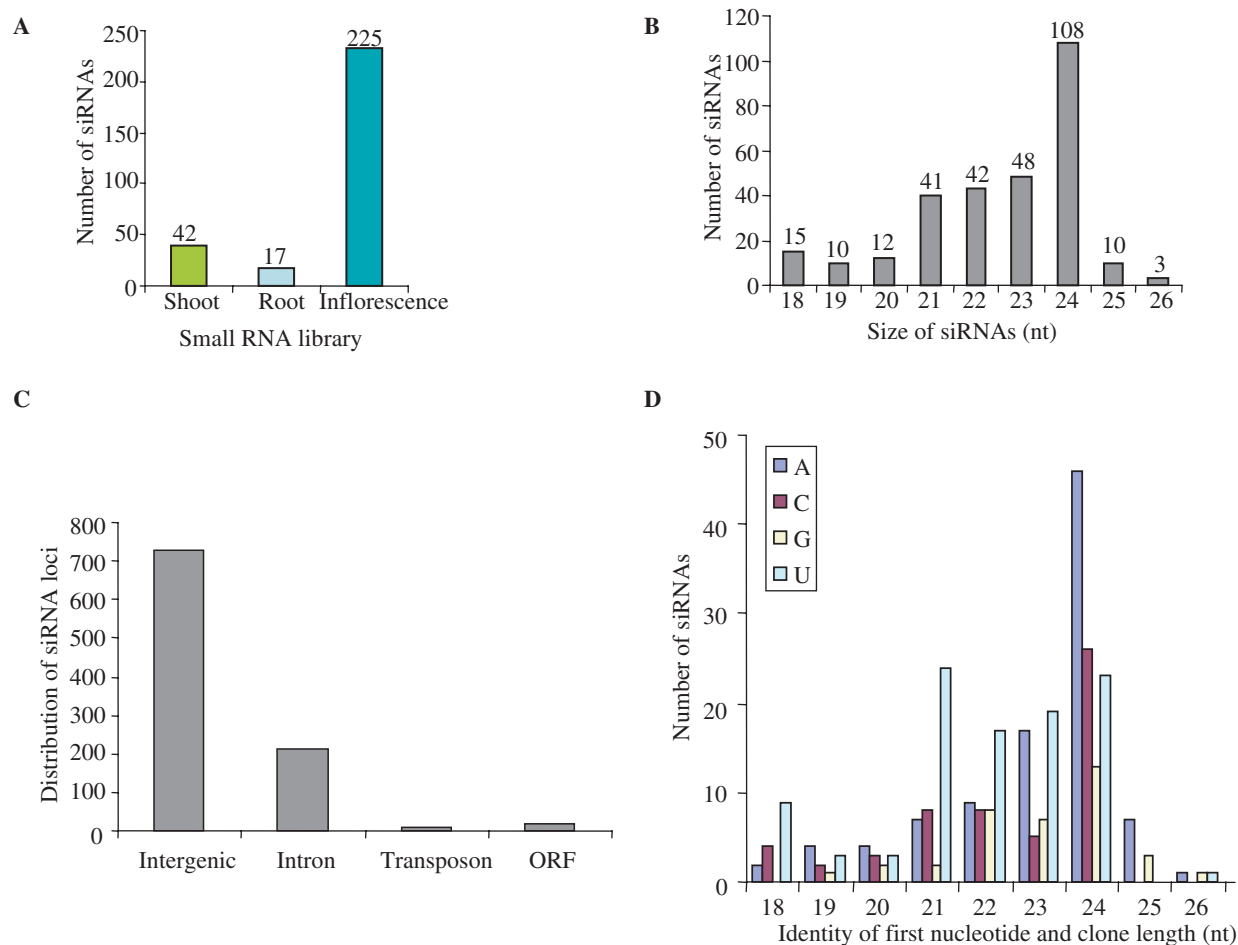
Figure 1. Rice endogenous siRNAs. (**A**) Histogram of the number of siRNAs identified from the libraries of shoots and roots of young seedlings, and inflorescence tissues. (**B**) Size distribution of endogenous siRNAs. (**C**) The rice endogenous siRNAs correspond to sequences from ORFs, transposons, introns and intergenic regions. (**D**) Sequence composition of the 5′ ends of rice siRNAs as a function of length.

21 and 24 nt size classes of plant siRNAs. Most of the endogenous siRNAs of the 21 nt class begin with a 5′ U, whereas the 5′ nt of the 21 nt class transgene siRNAs show an even distribution among A, U, C and G (13) (http://asrp. cgrb.oregonstate.edu/). The bias of the endogenous rice and *Arabidopsis* siRNAs or non-bias of the transgene siRNAs in their 5′ nt may reflect the specific activities of DCL involved in their biogenesis. Some of the putative siRNAs of the ~21 nt size class could be miRNAs. These could be generated from atypical hairpin structures that cannot be recognized as miRNA precursors. Analysis of the 24 nt size class in *Arabidopsis* further confirmed that this class of endogenous siRNAs has a 5′ A preference. This observation could suggest that one of the DCLs in rice and *Arabidopsis* is responsible for the generation of the 24 nt class of siRNAs preferentially with A on the 5′ end of the molecule. It was hypothesized that DCL3 and RNA-dependent RNA polymerase 2 (RDR2) generate endogenous siRNAs primarily of the large size (~24 nt) class in *Arabidopsis* (19). Similarly, the DCL3 ortholog in rice may be involved in the generation of the ~24 nt class of siRNAs. Further studies are required to correlate the generation of endogenous small RNAs with the specific DCL and RdRP activities in rice.

Genomic organization of endogenous siRNAs in rice

The genomic locations of rice endogenous siRNAs identified in the present study are shown in Supplementary Table 1. The 284 siRNAs map to 942 loci on the 12 rice chromosomes (Figure 1C and Supplementary Table 1). The discrepancy between the number of identified siRNAs (284) and the number of their corresponding genomic loci (942) is due to the fact that some siRNAs have multiple loci. Among the siRNAs, 165 are encoded by single copy loci in the rice genome, whereas the remaining siRNAs have multiple loci (2–21) (Supplementary Table 1). A total of 729 loci can be mapped to intergenic regions, and 213 correspond to introns (110 in the sense polarity and 103 in the antisense polarity relative to the genes) (Supplementary Table 1 and Figure 1C). They are scattered across all chromosomes with several islands of higher density (Supplementary Figure 1). The total number of siRNA loci between the different chromosomes is highly variable. The total number of siRNA loci peaks on chromosomes 11 and 12, whereas chromosome 5 has the lowest density and number of loci (Supplementary Figure 1). However, we were unable to detect a marked density correlation between siRNAs and genes or transposons along the different chromosomes.

Table 2. siRNAs that are antisense to protein-coding genes

siRNA	Accession number	Protein class	Position in gene
P1-C04	9638.m03532	Flavanone 3-hydroxylase	3′-UTR
	9632.m05823	Tubulin-folding co-factor B	3′-UTR
P7-E5	9640.m04164	Unknown protein	ORF
P8-G3	9636.m01529	DNA-directed RNA polymerase	ORF
P56-H4	9639.m01348; 9638.m03362; 9638.m01465	Unknown proteins	ORF
	9635.m02201; 9635.m01014; 9634.m01827	Unknown proteins	ORF
	9629.m04919; 9636.m02666; 9635.m02202	Unknown protein	ORF
	9633.m00440; 9631.m02402	Unknown protein	ORF
	9639.m03577	Leucine-rich repeat	ORF
P78-E8	9640.m00107	Unknown protein	ORF
	9640.m00111	Unknown protein	ORF
P83-D5	9640.m00279	Receptor protein kinase	ORF
P108-B7	9632.m02934	Unknown protein	ORF
P109-D12	9639.m00201	Auxin transport protein-like	ORF

Table 3. siRNAs originating from antisense orientations to transposons

siRNA Id	Accession number	Transposon annotation	Position in gene
P33-D12	9633.m01157	Gag-pol precursor	ORF
P40-H12	9633.m00737	Retrotransposon, Ty3-gypsy sub-class	ORF
P56-H4	9640.m00740; 9629.m02858; 9631.m04579	MuDR family transposases	ORF
	9632.m00690	Retrotransposon, Ty3-gypsy sub-class	ORF
	9640.m02844; 9638.m02247; 9638.m00643	Transposons, mariner sub-class	ORF
	9634.m00741; 9633.m03362; 9633.m02293	Transposons, mariner sub-class	ORF
	9631.m01776; 9629.m04738; 9629.m02757	Transposons, mariner sub-class	ORF
	9636.m02871; 9634.m02781;9632.m01127	Transposons, mariner sub-class	ORF
	9632.m02179	Transposons, mariner sub-class	ORF
P90-B8	9637.m00782; 9635.m02300	Retrotransposons, unclassified	ORF
	9635.m01634; 9633.m01299	Retrotransposons, Ty3-gypsy sub-class	ORF
P107-E1	9633.m01299; 9635.m01634	Retrotransposons, Ty3-gypsy	ORF

Some of the siRNAs display perfect complementarity with protein-coding genes and transposons (Tables 2 and 3). A total of 8 of the siRNAs (P1-C04, P7-E5, P8-G3, P56-H4, P78-E8, P83-D5, P108-B7 and P109-D12) are perfectly complementary to the mRNAs of 21 protein-coding genes (Table 2). Some of these siRNAs are complementary to multiple genes. For example, P56-H4 is complementary to 12 genes, and P1-C04 to 2. Similarly, 5 of the siRNAs (P33-D12, P40-H12, P56-H4, P90-B8 and P107-E1) are perfectly complementary to 25 of the transposons (Table 3). These include several classes of transposons (DNA transposons, retrotransposons and transposases). This observation implies that these siRNAs originate from the same locus as their respective target genes. However, some of these siRNAs appear to have multiple loci, and therefore it cannot be confirmed whether these siRNAs are derived from antisense transcription of the same locus or originate from a different locus and target the genes in *trans*. In summary, the endogenous siRNAs from rice largely originate from intergenic regions, but a significant proportion also arises from introns, exons and repetitive sequences (transposons and retroelements).

Rice siRNA clusters

siRNA loci often generate multiple, overlapping clusters of small RNAs. This stands in contrast to miRNA loci that generally yield a single miRNA. We have identified six clusters in intergenic regions, each with two or more siRNAs. These clustered siRNAs are spaced within an interval of 500 nt (Figure 2). Some of the clusters contain small RNAs that map to the plus and minus strands of short genomic regions (Figure 2). This observation suggests the possibility that the different siRNAs in a cluster originate from one long dsRNA.

Expression pattern of rice endogenous siRNAs

The tissue- and developmental stage-specific expression of small RNAs can provide clues about their physiological function. In *C.elegans*, the expression of several tncRNAs was found to depend on developmental stages (28). To examine whether rice endogenous siRNAs are expressed in a developmental stage-dependent and/or tissue-specific manner, we performed northern analysis using low molecular weight RNA blots from various tissues of mature plants (root, leaf and inflorescence) as well as 4-week-old young seedlings. For expression analysis, we tried to select siRNAs that appeared relatively frequently in our sequences or are represented by more loci in the genome, as these may have relatively high expression and thus more amenable for detection.

The northern analysis revealed that most of the 30 siRNAs tested are expressed at very low abundance and could not be detected on the small RNA blots. However, we could detect a signal at the appropriate sizes for six of the siRNAs (P16-C3, P4-D6, P88-A11, P98-A9, P108-A7 and P107-E1) (Figure 3). Two of the siRNAs are expressed relatively abundantly (P16-C3 and P107-E1). Some show tissue- or developmental
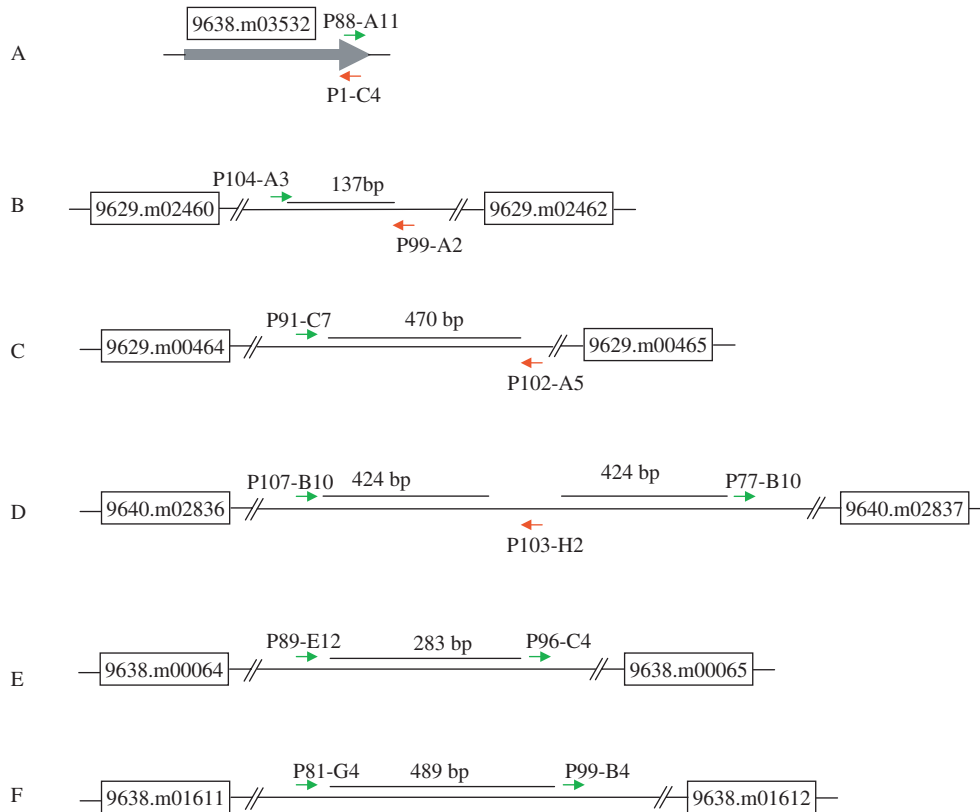
Figure 2. Endogenous siRNAs in clusters. The distance between two siRNAs is indicated. 5′ and 3′ flanking genes are shown. siRNAs are indicated by green or red arrows. For siRNAs that overlap the ORF of a gene, the siRNAs are shown on top and bottom of the gene. Location of siRNA on the genome not drawn to scale. A–F are different siRNA clusters.
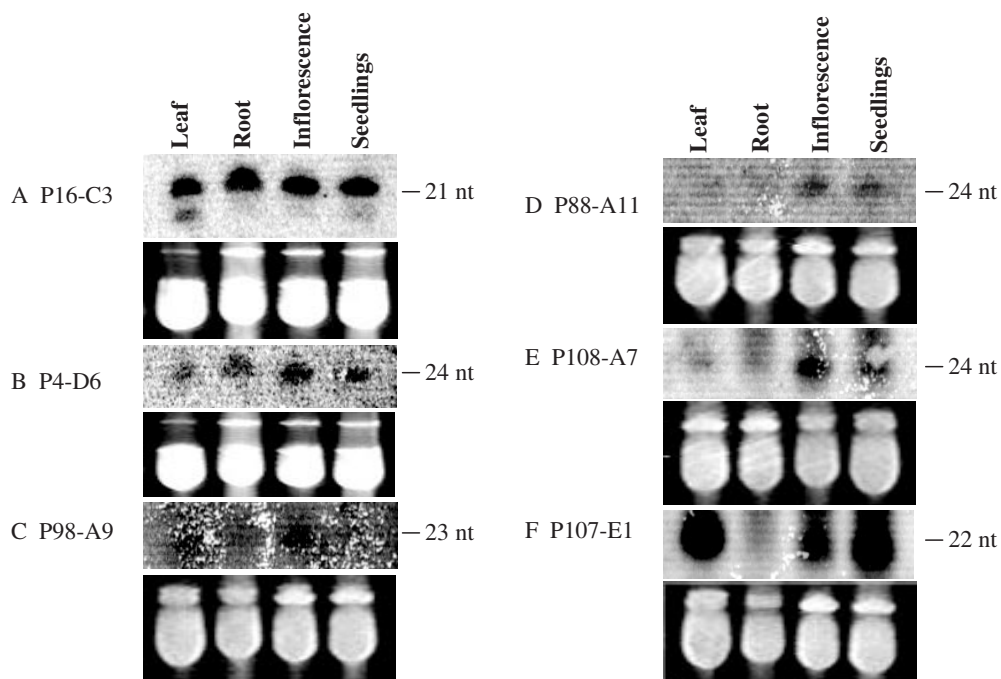
Figure 3. Expression analysis of some of the rice endogenous siRNAs by northern hybridization. Northern blots of low molecular weight RNA isolated from different tissues were probed with labeled oligonucleotides. The tRNA and 5S rRNA bands were visualized by ethidium bromide staining of polyacrylamide gels and served as loading controls. Labeled RNA oligonucleotide was used as size marker and the position was indicated. A–F are different siRNAs as listed.

Figure 4. PCR-based detection of endogenous siRNAs. PCR amplification was performed using a 5′ primer for the 5′ adapter sequence and a 3′ primer specific for the candidate siRNAs.

stage-specific expression patterns (P88-A11, P98-A9, P108-A7 and P107-E1). P16-C3 and P4-D6 appear to be uniformly expressed in all rice tissues examined (Figure 3A and B). In addition to the expected signal at 21 nt, the P16-C3 probe also detected a weaker signal at a lower size of ∼18–19 nt in all tissues examined. The lower sized signal may represent smaller sized siRNA rather than a d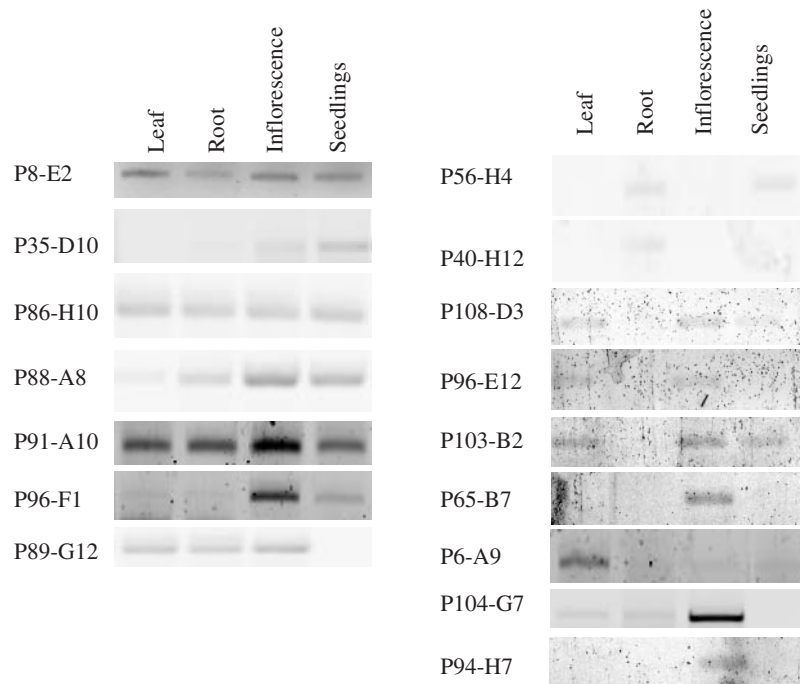egradation product of P16-C3. Furthermore, this small band was stronger in the leaf tissue than in other tissues. P16-C3 has multiple loci, and thus the smaller molecule might have originated from a different locus than the larger one. P98-A9 displayed moderate expression in leaves and young seedlings (Figure 3C). The expression of P88-A11 was moderate in inflorescence and seedlings but very weak in leaf and root tissues (Figure 3D). P108-A7 was expressed in all tissues tested, but the levels were higher in inflorescence and young seedlings (Figure 3E). The expression of P107-E1 was strong in leaves and young seedlings but almost below the detection limit in the root tissue (Figure 3F).

Furthermore, using an RT–PCR approach we were able to detect the expression of 16 additional siRNAs out of 65 tested (Figure 4). Some of these siRNAs, P8-E2, P86-H10, P88-A8, P91-A10, appear to be ubiquitously expressed, while others such as P108-D3, P103-B2, P65-B7, P104-G7 and P94-H11 are preferentially expressed in the inflorescence (Figure 4). The rest were expressed in at least one of the tissues tested (Figure 4). These results show that some endogenous siRNAs in rice are differentially expressed in different tissues and/or developmental stages.

Predicted targets

To identify potential *trans* mRNA targets for the endogenous siRNAs, we searched rice mRNAs for antisense hits with three or fewer mismatches and no mismatch was allowed between positions 10 and 11 from the 5′ end of the siRNA. *Arabidopsis* miRNA and tasiRNAs targets are known to exhibit up to three or even four mismatches to their target sites. The number of mismatches is typically lower in the 5′ region of the small RNA (16,17,37,38). By allowing a maximum of three mismatches, we were able to predict 111 *trans*-targets for 44 of the siRNAs (Table 4). Targets could not be predicted for the remaining siRNAs by applying these criteria. miRNA target sites in plants are usually found in the open reading frames (ORFs) of the target genes (15,39), although some target sites are in the UTR of target genes (15). Similarly, the endogenous tasiRNAs in *Arabidopsis* appear to have their target sites in the ORFs of the target genes (16,17). Most (87) of our predicted *trans*-target genes in rice have target sites in the ORFs. Eighteen have predicted target sites in 3′-UTRs and six in 5′-UTRs.

Some of the siRNAs are perfectly complementary to the protein-coding genes and transposable elements (Tables 2 and 3). These siRNAs are possibly involved in the *cis*-silencing of those protein-coding genes and transposons. In addition, some transposable elements were predicted as *trans*-targets for some of the siRNAs. A total of 51 of the 111 predicted *trans*-targets are transposons, and these are targeted by 16 siRNAs (Table 4). In all the cases, the target sites are located in ORFs of the transposons. These include mariner and En/Spm type transposons, retrotransposons (Ty3-gypsy, Ty1-copia), MuDR transposases and some other unclassified transposable elements.

The remaining predicted targets appear to have roles in a broad range of physiological processes (Table 4). These include genes encoding protein kinases, F-box proteins, Zn-finger proteins, disease-related proteins, lipase and another 45 proteins with unknown functions (Table 4). The predicted

Table 4. Predicted *trans*-targets of endogenous siRNAs from rice

siRNA	Predicted *trans*-target	ORF/UTR	protein class	MFGO
p1-C04	9639.m03698 (1); 9631.m01833 (1); 9633.m04353 (2); 9631.m03852 (2)	ORF	Unknown proteins	n.a.
P6-A9	9631.m05051 (2)	5′-UTR	Homeobox 1 protein OSH1	DNA binding
P8-C12	9635.m03442 (2)	ORF	Unknown protein	n.a.
P9-F11	9630.m04448 (1); 9636.m03634 (1); 9638.m01327 (1)	ORF	Unknown proteins	n.a.
	9632.m00040 (2)	ORF	Transposon	n.a.
	9634.m03300 (2)	ORF	Unknown protein	n.a.
P31-G9	9631.m02155 (1)	ORF	Protein kinase C-terminal domain	Kinase activity
P33-D12	9633.m01156 (1)	ORF	Unknown protein	n.a.
P40-H12	9633.m01657 (1); 9635.m01160 (1); 9640.m02569 (1); 9639.m02516 (1)	ORF	Retrotransposon proteins, Ty3-gypsy	n.a.
	9629.m02836 (2); 9630.m02869 (2); 9632.m01522 (2); 9633.m01553 (2)	ORF	Retrotransposon protein, Ty3-gypsy	n.a.
	9634.m01523 (2); 9638.m00819 (2); 9638.m01484 (2); 9639.mo2508 (2)	ORF	Retrotransposon protein, Ty3-gypsy	n.a.
	9632.m01381 (2); 9632.m05483 (2); 9633.m02072 (3); 9635.m00231 (3)	ORF	Retrotransposon protein, Ty3-gypsy	n.a.
	9636.m01098(3); 9633.m02207 (2); 9640.m01398 (2); 9635.m00348 (3)	ORF	Retrotransposon protein, Ty3-gypsy	n.a.
	9636.m02070 (3); 9633.m01607 (2)	ORF	Retrotransposon proteins, unclassified	n.a.
	9636.m01059 (2)	ORF	Transposon protein, CACTA, En/Spm	n.a.
P49-E4	9632.m05379 (2)	5′-UTR	G10 protein	Nucleus
P56-H4	9635.m04625 (1); 9637.m00162 (1); 9635.m04145 (1); 9629.m01794 (1)	ORF	Transposon proteins, mariner	n.a.
	9630.m02898 (1)	ORF	MuDR family transposases	n.a.
	9635.m02910 (1); 9635.m04732 (1)	ORF	Zinc finger proteins	n.a.
	9631.m00145 (1); 9631.m04466 (1)	ORF	Unknown proteins	n.a.
P71-G7	9636.m03978 (1)	ORF	b 1,2-xylosyltransferase	n.a.
	9629.m01388 (1)	5′-UTR	BT1 family	Transporter activity
P75-D3	9632.m00807 (1)	ORF	Retrotransposon protein, Ty1-copia	n.a.
P77-H6	9631.m03009 (2); 9632.m02240 (2); 9629.m02778 (2); 9629.m02218 (2)	ORF	Retrotransposon proteins, Ty3-gypsy	n.a.
	9631.m03445 (2)	ORF	Polyprotein	n.a.
P79-H4	9639.m02976 (1)	3′-UTR	Unknown protein	n.a.
	9638.m03532 (1)	3′-UTR	Flavanone 3-hydroxylase	Catalytic activity
	9632.m05823 (1)	3′-UTR	Tubulin folding co-factor B	Cell organization and biogenesis
	9631.m01833 (2)	3′-UTR	Ribosomal protein S7e	Structural molecule activity
	9639.m03698 (2); 9633.m04353 (3)	ORF	Unknown proteins	n.a.
P81-B6	9630.m05360 (2)	3′-UTR	Unknown protein	n.a.
P81-F4	9630.m00799 (2)	3′-UTR	Chorismate mutase	Catalytic activity
P83-G5	9632.m04071 (1)	3′-UTR	Polyadenylate binding protein family	RNA binding
P84-C2	9634.m04235 (2)	ORF	Unknown proteins	n.a.
P86-F9	9640.m03178 (2); 9639.m01190 (2); 9637.m00219 (2); 9635.m00791 (2)	ORF	Unknown proteins	n.a.
	9634.m02637 (2)	ORF	Retrotransposon protein, Ty1-copia	n.a.
P86-G10	9637.m01993 (2)	5′-UTR	Unknown protein	n.a.
P86-H10	9636.m00899 (3)	ORF	N7 like-protein	Catalytic activity
P87-A3	9634.m02242 (2)	ORF	Unknown protein	n.a.
	9633.m01814 (2)	ORF	Transposon protein, CACTA, En/Spm	n.a.
P87-H6	9640.m01018 (2)	ORF	Transposon protein, CACTA, En/Spm	n.a.
P88-A11	9630.m00104 (2)	5′-UTR	Probable protein kinase	Kinase activity
	9634.m00364 (3)	3′-UTR	Unknown protein	n.a.
P88-H6	9629.m01127 (1)	3′-UTR	Lipase/acylhydrolase	Hydrolase activity
P89-B12	9629.m01463 (3)	ORF	Unknown protein	n.a.
P89-G12	9639.m01787 (2); 9634.m02818 (2); 9634.m02252 (3)	ORF	Transposon proteins, CACTA, En/Spm	n.a.
P89-H4	9629.m01059 (2)	3′-UTR	WH2 motif containing protein	n.a.
	9634.m02256 (2)	5′-UTR	Similar to Xet3 protein	Hydrolase activity
P91-E4	9640.m02704 (2)	3′-UTR	MLA12	Response to biotic stimuli
P92-A9	9631.m00634 (2)	ORF	Retrotransposon protein, unclassified	n.a.
P94-D12	9631.m04248 (3)	3′-UTR	Unknown protein	n.a.
P96-E12	9635.m01047 (3)	ORF	Cellulose synthase	Transferase activity
	9629.m00460 (2)	ORF	Retrotransposon protein, Ty1-copia	n.a.
P98-A9	9636.m02299 (1)	ORF	Unknown protein	n.a.
	9633.m01348 (2)	ORF	Retrotransposon protein, Ty3-gypsy	n.a.
P98-G5	9632.m01377 (3)	ORF	BURP domain-containing protein	Hydrolase activity
	9630.m01508 (3)	ORF	Xa1-like (bacterial blight resistance protein)	Response to biotic stimulus
P102-C12	9638.m00293 (1)	3′-UTR	F-box protein	n.a.
	9629.m07044 (1)	3′-UTR	Unknown protein	n.a.

Table 4. *Continued*

siRNA	Predicted *trans*-target	ORF/UTR	protein class	MFGO
P102-F3	9629.m00936 (1)	3′-UTR	Zinc finger, C2H2 type	Nucleic acid binding
P105-G11	9630.m04940 (2); 9636.m01694 (2)	ORF	Unknown proteins	n.a.
P106-F8	9632.m02318 (2)	3′-UTR	Unknown protein	n.a.
P107-E1	9637.m00211 (1); 9633.m01405 (2); 9634.m02637 (2)	ORF	Retrotransposon proteins, Ty3-gypsy	n.a.
P107-E2	9633.m03971 (2)	ORF	Unknown protein	n.a.
	9633.m03923 (3)	3′-UTR	Unknown protein	n.a.
P107-H3	9629.m04373 (2)	ORF	Retrotransposon protein, unclassified	n.a.
	9633.m03827 (2)	ORF	Unknown protein	n.a.
P108-A7	9639.m02094 (2)	ORF	Retrotransposon protein, Ty3-gypsy	n.a.
	9639.m03271 (2)	ORF	Transposon protein, CACTA, En/Spm	n.a.
P108-H5	9638.m02410 (2)	ORF	Unknown protein	n.a.
P108-H10	9629.m05975 (2)	ORF	Transposon protein, unclassified	n.a.
P109-E3	9631.m01773 (2)	3′-UTR	Leucoanthocyanidin dioxygenase	Catalytic activity

MFGO, molecular function gene ontology; n.a., not available.

Figure 5. Identification of siRNA-guided cleavage products of target mRNAs in rice. (**A**) mRNA 9639.m00201, (**B**) mRNA 9636.m02299 and (**C**) mRNA 9632.m00807. Mapping of cleavage sites was performed by RLM-5′ RACE. Partial mRNA sequences from target genes were aligned with siRNAs. Numbers indicate the fraction of cloned PCR products terminating at different positions.

target of siRNA 109-D12 is 9629.m00201, a PIN1-like gene, which has two complementary sites within the ORF (Figure 5A). The two target sites correspond to positions 362–383 and 1397–1418, with zero mismatches (Figure 5A). P109-D12 has two genomic loci corresponding to the two target sites in 9629.m00201, which indicates that P109-D12 siRNA originates possibly by antisense transcription.

P1-C04 and P79-H4 are of the same size (24 nt) and differs by 1 nt and the origins were different. P1C04 is perfectly complementary to the 3′-UTR of flavanone 3-hydroxylase. P79-H4 is predicted to target the same site but with a mismatch at the position 11 from the 5′ end of the siRNA.

Endogenous rice siRNAs can direct the cleavage of target mRNAs

The cleavage of target mRNAs appears to be the predominant mode of gene regulation of miRNAs in plants. In addition to miRNAs, tasiRNAs in *Arabidopsis* have also been shown to direct the cleavage of the predicted target transcripts (16,17). Target mRNA fragments resulting from siRNA-guided cleavage are characterized by having a 5′ phosphate group, and cleavage occurs near the middle of the base pair interaction region. To examine whether some of the predicted siRNA targets are true targets and whether these siRNAs can direct the cleavage of predicted targets, we used a RNA ligase-mediated 5′ RACE procedure to map the cleavage site of siRNA targets. Using this procedure, we were able to detect and clone the cleaved fragments of 9629.m00201 (PIN1-like gene) targeted by P109-D12 (Figure 5A). Interestingly, this gene has two target sites in its ORF, and the cleaved fragments corresponded to both of the target sites. This observation provides an example of *cis*-silencing by endogenous siRNAs. We have also found evidence for siRNA-mediated *trans*-targeting in rice. For example, P98-A9 targets the unknown protein 9636.m02299 and P75-D3 targets the retrotransposon

9632.m00807. Both are examples of *trans*-targeting by siRNAs (Figure 5B and C). In both the cases, the most common 5′ end of the mRNA fragment is mapped to the nucleotide that pairs to the 10th nucleotide of the siRNA from its 5′ end. In this study, a substantial number of endogenous siRNAs were predicted to target protein-coding transcripts, and even if a small portion of these transcripts are genuine targets, the potential of siRNA regulation could be vast.

DISCUSSION

Endogenous siRNAs in plants can be divided into two classes on the basis of size and function: the 21 nt siRNAs that direct post-transcriptional silencing via mRNA degradation and the 24 nt siRNAs that trigger methylation of homologous DNA leading to TGS (10). Endogenous siRNAs in rice may be grouped into similar size and functional categories. A total of 284 unique putative siRNA sequences corresponding to 942 genomic loci have been identified in this study. The expression of some of the siRNAs is tissue specific or developmental stage specific. We predicted a large number of mRNA targets for the siRNAs. Importantly, we validated three of the predicted targets and provided evidence for both *cis*-silencing and *trans*-silencing of target mRNAs by rice siRNA-guided cleavage. Most of the siRNAs (225 of 284) were identified from an inflorescence library. The reason for the high number of endogenous siRNAs from inflorescence tissues is unknown, but may implicate an involvement of siRNAs in active meristematic/cell division or developmental processes. A large number of the siRNA loci (729) map to intergenic regions. It is unknown how the siRNAs mapped to intergenic regions are regulated, but they likely have their own regulatory sequences. Our results suggest that endogenous siRNAs in rice is very diverse.

The endogenous siRNAs differ in their 5′ end nucleotide preference from NOS promoter-driven transgenic siRNAs. *Arabidopsis* NOS promoter-expressing plants generate siRNAs of both 21 and 24 nt size classes (13). Two major differences were observed with respect to preferences for 5′ nt in rice endogenous siRNAs. Most of the endogenous siRNAs of the 21 nt size class begin with 5′ U, whereas 5′ nt of the 21 nt size class from transgenic siRNAs show an even distribution in nucleotides (13). The preference in the rice endogenous siRNAs could reflect a specific DCL activity in rice. Some siRNAs of this size class could still be miRNAs but cannot be defined as such, because they may be generated from atypical hairpin structures. With respect to the 24 nt size class, 5′ A was the predominant nucleotide for the endogenous siRNAs in rice, whereas C was the major 5′ nt in transgene-derived siRNAs. Although the 5′ A bias for the 24 nt size class was known from previous findings in *Arabidopsis* (9), our study confirms this striking difference *in vivo* in rice. It is interesting that this distinct feature is not apparent in the 24 nt size class of endogenous siRNAs from the lower plant moss (*Polytrichum juniperinum*) (40). This observation raises the possibility that the DCL responsible for processing the 24 nt size class with 5′ adenine may be absent in lower plants. Further studies with individual DCL knock-out mutants may verify the role of different DCLs in conferring the 5′ nt preference of endogenous siRNAs in rice.

The expression of several tncRNAs from *C.elegans* was shown to be constitutive and ubiquitous, whereas many other tncRNAs exhibit preferential expression in a temporal or tissue-specific manner (28). Our northern and RT–PCR results suggest that some of the rice endogenous siRNAs are expressed in a tissue-specific and/or developmental stage-dependent manner. The differential expression pattern supports a regulatory role of these siRNAs in specific tissues or in development.

Many siRNAs are expected to act on the same locus from which they originate. Presumably, these siRNAs are derived from the same loci as their targets, through antisense transcription or RdRP activity. Some of the rice endogenous siRNAs originate from the antisense strands of protein- or transposon-coding genes and might mediate the post-transcriptional silencing of these genes.

We also predicted 111 genes as potential *trans*-targets of 44 endogenous siRNAs in rice. Transposons represent ∼46% of the predicted *trans*-targets of siRNAs in the present study. Our analysis suggests that some of the retroelements that are predicted targets of siRNAs are whole. For example, out of the 31 gypsy type of retrotransposons that were predicted targets of siRNAs, 13 appear to have all constituent proteins without nonsense mutations, but the remaining 18 are truncated. Similarly, one out of the four CACTA retrotransposons that were predicted targets of siRNAs appears to be whole, whereas the remaining three are truncated. Small RNAs corresponding to transposon sequences have also been detected in *Arabidopsis* (41,42). Transposable elements are DNA sequences that can move and multiply within the genome of an organism. Higher eukaryotic genomes contain a large number of transposable elements, e.g. 45% of the human genome consists of remnants of transposon sequences. Twelve percent of the *C.elegans* genome is transposon sequences. Transposable elements make up 14% of the *Arabidopsis* genome, in contrast to an estimated 50–80% of the maize genome (43,44). The rice genome is populated by representatives from all known transposon superfamilies, including elements that cannot be easily classified into either Class I or II (45). Present estimates of the transposon content of the *O.sativa* ssp. *Japonica* genome is at least 35%. Transposon activation can have a range of effects, including alterations in gene expression, gene deletion and insertion, and chromosome rearrangements, most of which are deleterious. Recent work in plants and other organisms has revealed that both TGS and PTGS are important for the silencing of transposons. In organisms other than plants, PTGS is likely to be the primary pathway of transposon silencing (46). Because siRNAs guide the sequence-specific RNA degradation that occurs during RNAi, transposon siRNAs might mediate the degradation of transposon mRNAs to establish transposon silencing. rasiRNAs, which are derived from every known type of transposon in the *Drosophila* genome, could potentially regulate transposon mobility and assembly of heterochromatin at transposon-rich regions, such as telomeres and centromeres (27). In addition to the endogenous siRNAs that correspond to transposons in antisense orientations, our analysis predicted transposons as likely *trans*-targets for several of the siRNAs. This result implies a role for endogenous siRNAs in controlling the proliferation of transposons in both *auto*- and *trans*-silencing manners in rice.

The remaining predicted targets appear to have roles in diverse physiological processes and include both regulatory as well as metabolism and cell structure-related genes (Table 4). In addition, 45 unknown proteins are also predicted targets for some of the siRNAs. The validated targets in this study include 9629.m00201, which serves as an example for *cis*-silencing by siRNA P109-D12. Two other validated targets are examples of *trans*-targeting by siRNAs. Thus, our analysis suggests that endogenous siRNAs in rice can guide mRNA cleavage both in *cis* and in *trans*. Many of the siRNAs that do not have predicted mRNA targets may target DNA for TGS in rice.

The profile of naturally occurring siRNAs in rice provides an important foundation to explore the potential roles of these molecules in genome maintenance, genome expression and defense. The large number of predicted targets and our experimental evidence showing *cis* and *trans* silencing of genes by siRNA-guided mRNA cleavage suggest that endogenous siRNAs are important for the regulation of many genes in rice.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at NAR Online.

ACKNOWLEDGEMENTS

REFERENCES

1. Carrington,J.C. and Ambros,V. (2003) Role of microRNAs in plant and animal development. *Science*, **301**, 336–338.
2. Ambros,V. (2004) The functions of animal microRNAs. *Nature*, **431**, 350–355.
3. Bartel,D.P. (2004) MicroRNAs: genomics, biogenesis, mechanism and function. *Cell*, **116**, 281–297.
4. Baulcombe,D. (2004) RNA silencing in plants. *Nature*, **431**, 356–363.
5. Hamilton,A.J. and Baulcombe,D.C. (1999) A species of small antisense RNA in posttranscriptional gene silencing in plants. *Science*, **286**, 950–952.
6. Hammond,S.M., Bernstein,E., Beach,D. and Hannon,G.J. (2000) An RNA-directed nuclease mediates post-transcriptional gene silencing in *Drosophila* cells. *Nature*, **404**, 293–296.
7. Zamore,P.D., Tuschl,T., Sharp,P.A. and Bartel,D.P. (2000) RNAi: double-stranded RNA directs the ATP-dependent cleavage of mRNA at 21 to 23 nucleotide intervals. *Cell*, **101**, 25–33.
8. Billy,E., Brondani,V., Zhang,H., Muller,U. and Filipowicz,W. (2001) Specific interference with gene expression induced by long, double-stranded RNA in mouse embryonal teratocarcinoma cell lines. *Proc. Natl Acad. Sci. USA*, **98**, 14428–14433.
9. Tang,G., Reinhart,B.J., Bartel,D.P. and Zamore,P.D. (2003) A biochemical framework for RNA silencing in plants. *Genes Dev.*, **17**, 49–63.
10. Hamilton,A., Voinnet,O., Chappell,L. and Baulcombe,D. (2002) Two classes of short interfering RNA in RNA silencing. *EMBO J.*, **21**, 4671–4679.
11. Mallory,A., Reinahrt,B., Bartel,D., Vance,V. and Bowman,L. (2002) A viral suppressor of RNA silencing differentially regulates the accumulation of short interfering RNAs and micro-RNAs in tobacco. *Proc. Natl Acad. Sci. USA*, **99**, 15228–15233.
12. Llave,C., Kasschau,K.D., Rector,M. and Carrington,J.C. (2002) Endogenous and silencing-associated small RNAs in plants. *Plant Cell*, **14**, 1605–1619.
13. Papp,I., Mette,M.F., Aufsatz,W., Daxinger,L., Schauer,S.E., Ray,A., van der Winden,J., Matzke,M. and Matzke,A.J.M. (2003) Evidence for nuclear processing of plant micro RNA and short interfering RNA precursors. *Plant Physiol.*, **132**, 1382–1390.
14. Nicolás,F.E., Torres-Martínez,S. and Ruiz-Vázquez,R.M. (2003) Two classes of small antisense RNAs in fungal RNA silencing triggered by non-integrative transgenes. *EMBO J.*, **22**, 3983–3991.
15. Sunkar,R. and Zhu,J.K. (2004) Novel and stress regulated microRNAs and other small RNAs from *Arabidopsis. Plant Cell*, **16**, 2001–2019.
16. Vazquez,F., Vaucheret,H., Rajagopalan,R., Lepers,C., Gasciolli,V., Mallory,A.C., Hilbert,J.L., Bartel,D.P. and Crete,P. (2004) Endogenous *trans*-acting siRNAs regulate the accumulation of *Arabidopsis* mRNAs. *Mol. Cell*, **16**, 69–79.
17. Peragine,A., Yoshikawa,M., Wu,G., Albrecht,H.L. and Poethig,R.S. (2004) *SGS3* and *SGS2/SDE1/RDR6* are required for juvenile development and the production of *trans*-acting siRNAs in *Arabidopsis. Genes Dev.*, **18**, 2368–2379.
18. Zilberman,D., Cao,X. and Jacobsen,S.E. (2003) ARGONAUTE4 control of locus-specific siRNA accumulation and DNA and histone methylation. *Science*, **299**, 716–719.
19. Xie,Z., Kasschau,K.D. and Carrington,J.C. (2003) Negative feedback regulation of *Dicer-Like1* in *Arabidopsis* by microRNA-guided mRNA degradation. *Curr. Biol.*, **13**, 784–789.
20. Chan,S.W., Zilberman,D., Xie,Z., Johansen,L.K., Carrington,J.C. and Jacobsen,S.E. (2004) RNA silencing genes control *de novo* DNA methylation. *Science*, **303**, 1336.
21. Park,W., Li,J., Song,R., Messing,J. and Chen,X. (2002) CARPEL FACTORY, a Dicer homolog, and HEN1, a novel protein, act in microRNA metabolism in *Arabidopsis thaliana. Curr. Biol.*, **12**, 1484–1495.
22. Reinhart,B.J., Weinstein,E.G., Rhoades,M.W., Bartel,B. and Bartel,D.P. (2002) MicroRNAs in plants. *Genes Dev.*, **16**, 1616–1626.
23. Kurihara,Y. and Watanabe,Y. (2004) *Arabidopsis* micro-RNA biogenesis through Dicer-like 1 protein functions. *Proc. Natl Acad. Sci. USA*, **101**, 12753–12758.
24. Boutet,S., Vazquez,F., Liu,J., Béclin,C., Fagard,M., Gratias,A., Morel,J.B., Crété,P., Chen,X. and Vaucheret,H. (2003) *Arabidopsis HEN1*: a genetic link between endogenous miRNA controlling development and siRNA controlling transgene silencing and virus resistance. *Curr. Biol.*, **13**, 843–848.
25. Onodera,Y., Haag,J.R., Ream,T., Nunes,P.C., Pontes,O. and Pikaard,C.S. (2005) Plant nuclear RNA polymerase IV mediates siRNA and DNA methylation-dependent heterochromatin formation. *Cell*, **120**, 613–622.
26. Herr,A.J., Jensen,M.B., Dalmay,T. and Baulcombe,D.C. (2005) RNA polymerase IV directs silencing of endogenous DNA. *Science*, **308**, 118–120.
27. Aravin,A., Lagos-Quintana,M., Yalcin,A., Zavalon,M., Marks,D., Snyder,B., Gaasterland,T., Meyer,J. and Tuschl,T. (2003) The small RNA profile during *Drosophila melanogaster* development. *Dev. Cell*, **5**, 337–350.
28. Ambros,V., Lee,R.C., Lavanway,A., Williams,P.T. and Jewell,D. (2003) MicroRNAs and other tiny endogenous RNAs in *C.elegans. Curr. Biol.*, **13**, 807–818.
29. Reinhart,B.J. and Bartel,D.P. (2002) Small RNAs correspond to centromere heterochromatic repeats. *Science*, **297**, 1831.
30. Ewing,P. and Green,B. (1998) Basecalling of automated sequencer traces using phred. II. Error probabilities. *Genome Res.*, **8**, 186–194.
31. Griffiths-Jones,S., Bateman,A., Marshall,M., Khanna,A. and Eddy,S.R. (2003) Rfam: an RNA family database. *Nucleic Acids Res.*, **31**, 439–441.
32. Altschul,S.F., Madden,T.L., Schaffer,A.A., Zhang,J., Zhang,Z., Miller,W. and Lipman,D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, **25**, 3389–3402.
33. Sunkar,R., Girke,T., Jain,P.K. and Zhu,J.K. (2005) Cloning and characterization of microRNAs from rice. *Plant Cell*, **17**, 1397–1411.
34. Dsouza,M., Larsen,N. and Overbeek,R. (1997) Searching for patterns in genomic data. *Trends Genet.*, **13**, 497–498.

35. Rice,P., Longden,I. and Bleasby,A. (2000) EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet.*, **16**, 276–277.

36. Grad,Y., Aach,J., Hayes,G.D., Reinhart,B.J., Church,G.M., Ruvkun,G. and Kim,J. (2003) Computational and experimental identification of *C.elegans* microRNAs. *Mol. Cell*, **11**, 1253–1263.

37. Palatnik,J.F., Allen,E., Wu,X., Schommer,C., Schwab,R., Carrington,J.C. and Weigel,D. (2003) Control of leaf morphogenesis by microRNAs. *Nature*, **425**, 257–263.

38. Jones-Rhoades,M.W. and Bartel,D.P. (2004) Computational identification of plant microRNAs and their targets, including a stress-induced miRNA. *Mol. Cell*, **14**, 787–799.

39. Rhoades,M.W., Reinhart,B.J., Lim,L.P., Burge,B.C., Bartel,B. and Bartel,D.P. (2002) Prediction of plant microRNA targets. *Cell*, **110**, 513–520.

40. Axtell,M.J. and Bartel,D.P. (2005) Antiquity of microRNAS and their targets in land plants. *Plant Cell*, **17**, 1658–1673.

41. Liu,J., He,Y., Amasino,R. and Chen,X. (2004) siRNAs targeting an intronic transposon in the regulation of natural flowering behavior in *Arabidopsis*. *Genes Dev.*, **18**, 2873–2878.

42. Lippman,Z., May,B., Yordan,C., Singer,T. and Martienssen,R. (2003) Distinct mechanisms determine transposon inheritance and methylation via small interfering RNA and histone modification. *PLoS Biol.*, **1**, E67.

43. Arabidopsis Genome Intiative (2000), Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature*, **408**, 796–815.

44. SanMiguel,P. and Bennetzen,J.L. (1998) Evidence that a recent increase in maize genome size was caused by the massive amplification of intergene retrotransposons. *Ann. Bot.*, **82**, 37–44.

45. Temnykh,S., De Clerck,G., Lukashova,A., Lipovich,L., Cartinhour,S. and Mc Couch,S. (2001) Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): frequency, length variation, transposon associations, and genetic marker potential. *Genome Res.*, **11**, 1441–1452.

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

47. Schramke,V. and Allshire,R. (2003) Hairpin RNAs and retrotransposon LTRs effect RNAi and chromatin-based gene silencing. *Science*, **301**, 1069–1074.