Molecular Cell, Vol. 14, 787–799, June 18, 2004, Copyright ©2004 by Cell Press

Computational Identification of Plant MicroRNAs and Their Targets, Including a Stress-Induced miRNA

Matthew W. Jones-Rhoades and David P. Bartel*
Whitehead Institute for Biomedical Research
and Department of Biology
Massachusetts Institute of Technology
9 Cambridge Center
Cambridge, Massachusetts 02142

Summary

MicroRNAs (miRNAs) are ~21-nucleotide RNAs, some of which have been shown to play important gene-regulatory roles during plant development. We developed comparative genomic approaches to systematically identify both miRNAs and their targets that are conserved in *Arabidopsis thaliana* and rice (*Oryza sativa*). Twenty-three miRNA candidates, representing seven newly identified gene families, were experimentally validated in *Arabidopsis*, bringing the total number of reported miRNA genes to 92, representing 22 families. Nineteen newly identified target candidates were confirmed by detecting mRNA fragments diagnostic of miRNA-directed cleavage in plants. Overall, plant miRNAs have a strong propensity to target genes controlling development, particularly those of transcription factors and F-box proteins. However, plant miRNAs have conserved regulatory functions extending beyond development, in that they also target superoxide dismutases, laccases, and ATP sulfurylases. The expression of miR395, the sulfurylase-targeting miRNA, increases upon sulfate starvation, showing that miRNAs can be induced by environmental stress.

Introduction

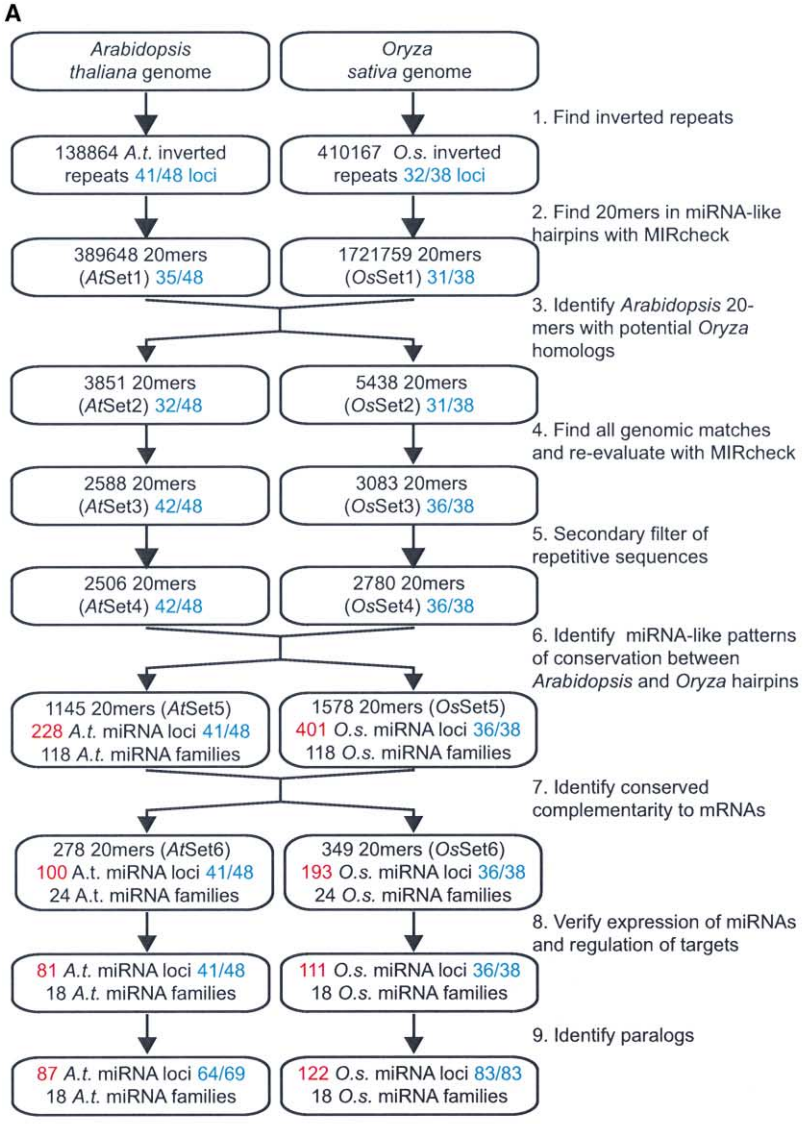
MicroRNAs are endogenous 20 to 24 nucleotide RNAs, some of which are known to play important posttranscriptional regulatory roles in plants and animals (Bartel and Bartel, 2003; Lai, 2003; Bartel, 2004). MicroRNAs are initially transcribed as much longer RNAs that contain imperfect hairpins, from which the mature miRNAs are excised by Dicer-like enzymes (Grishok et al., 2001; Hutvagner et al., 2001; Ketting et al., 2001; Lee et al., 2002, 2003; Park et al., 2002; Reinhart et al., 2002). Each mature miRNA derives from the double-stranded portion of the hairpin and is initially excised as a duplex comprising two ~22 nt RNAs, one of which is the mature miRNA while the other, known as the miRNA*, comes from the opposite arm of the hairpin (Lau et al., 2001; Reinhart et al., 2002; Khvorova et al., 2003; Lim et al., 2003b; Schwarz et al., 2003). The miRNA of this miRNA:miRNA* duplex is preferentially loaded into the RNA-induced silencing complex (RISC, Hammond et al., 2000), where it functions as a guide RNA to direct the posttranscriptional repression of mRNA targets, while the miRNA* is degraded (Hutvagner and Zamore, 2002; Mourelatos et al., 2002; Khvorova et al., 2003; Schwarz et al., 2003).

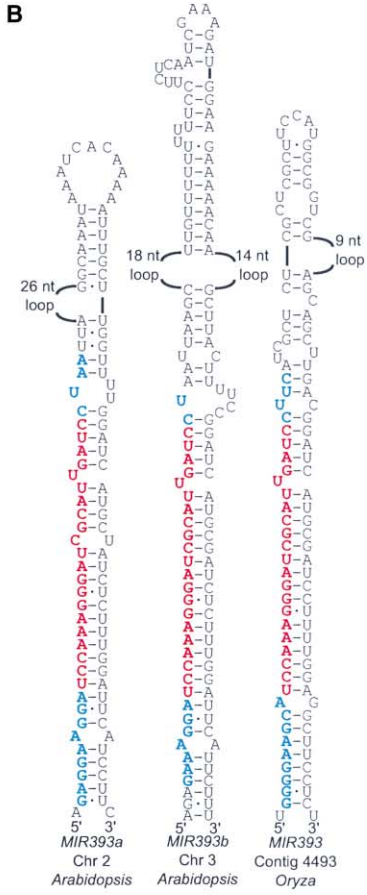
The primary method of identifying miRNA genes has been to isolate, reverse transcribe, clone, and sequence small cellular RNAs (Lagos-Quintana et al., 2001; Lau et al., 2001; Lee and Ambros, 2001; Llave et al., 2002a; Park et al., 2002; Reinhart et al., 2002). However, molecular cloning is biased toward finding miRNAs that are relatively abundant. In animals, miRNA gene discovery by molecular cloning has been supplemented by systematic computational approaches that identify evolutionarily conserved miRNA genes by searching for patterns of sequence and secondary structure conservation that are characteristic of metazoan miRNA hairpin precursors (Ambros et al., 2003b; Grad et al., 2003; Lai et al., 2003; Lim et al., 2003a, 2003b). The most sensitive of these methods indicate that miRNAs constitute nearly 1% of all predicted genes in nematodes, flies, and mammals (Lai et al., 2003; Lim et al., 2003a, 2003b). Methods developed in one animal lineage work well when extended to another animal lineage (Lim et al., 2003a), but cannot be directly applied to plants because the hairpins of plant miRNAs are more heterogeneous than those of animal miRNAs (Reinhart et al., 2002).

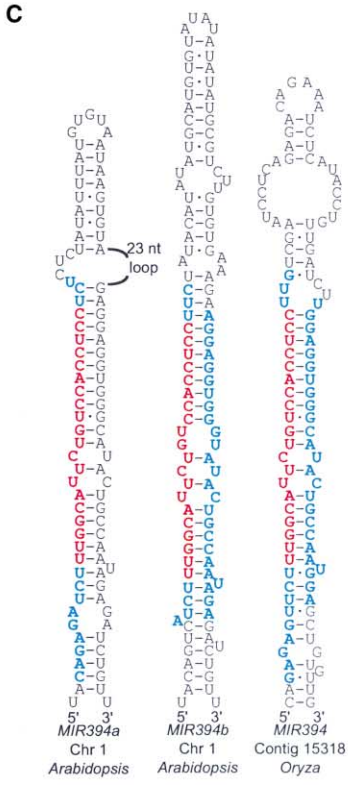
Because the miRNAs recognize their regulatory targets through base pairing, computational methods have been invaluable for identifying these targets. The extensive complementarity between plant miRNAs and mRNAs makes systematic target identification easier in plants than in animals (Rhoades et al., 2002). A search for targets of 13 *Arabidopsis* miRNA families predicted 49 unique targets, with a signal-to-noise ratio exceeding 10:1, simply by looking for *Arabidopsis* messages with three or fewer mismatches (Rhoades et al., 2002). Evolutionary conservation of the miRNA:mRNA pairing in rice (Rhoades et al., 2002), together with experimental evidence showing that these miRNAs direct cleavage of their predicted mRNA targets (Llave et al., 2002b; Kasschau et al., 2003; Palatnik et al., 2003; Tang et al., 2003; Mallory et al., 2004; Vazquez et al., 2004), supports the validity of these predictions. Because metazoan miRNAs only rarely recognize their targets with such extensive complementarity (Yekta et al., 2004), more sophisticated methods that search for short segments of conserved complementarity to the miRNAs are required to identify metazoan miRNA targets (Enright et al., 2003; Lewis et al., 2003; Stark et al., 2003).

The previously identified plant miRNAs have a remarkable propensity to target genes involved in development, particularly those of transcription factors (Rhoades et al., 2002). In all cases where disruption of plant miRNA regulation has been reported, striking developmental abnormalities are observed. Dominant gain-of-function mutations in HD-ZIP transcription factor genes *PHABULOSA*, *PHAVULOTA*, and *REVOLUTA* that destabilize pairing to miR165/miR166 cause loss of adaxial/abaxial polarity in developing leaves (McConnell et al., 2001; Rhoades et al., 2002; Emery et al., 2003; Kidner and Martienssen, 2004). In maize, similar mutations in the HD-ZIP gene *ROLLED LEAF1* also cause adaxilization of the abaxial

*Correspondence: dbartel@wi.mit.edu


A
Arabidopsis thaliana genome
Oryza sativa genome
1. Find inverted repeats
138864 A.t. inverted repeats 41/48 loci
410167 O.s. inverted repeats 32/38 loci
2. Find 20mers in miRNA-like hairpins with MIRcheck
389648 20mers (AtSet1) 35/48
1721759 20mers (OsSet1) 31/38
3. Identify Arabidopsis 20-mers with potential Oryza homologs
3851 20mers (AtSet2) 32/48
5438 20mers (OsSet2) 31/38
4. Find all genomic matches and re-evaluate with MIRcheck
2588 20mers (AtSet3) 42/48
3083 20mers (OsSet3) 36/38
5. Secondary filter of repetitive sequences
2506 20mers (AtSet4) 42/48
2780 20mers (OsSet4) 36/38
6. Identify miRNA-like patterns of conservation between Arabidopsis and Oryza hairpins
1145 20mers (AtSet5)
228 A.t. miRNA loci 41/48
118 A.t. miRNA families
1578 20mers (OsSet5)
401 O.s. miRNA loci 36/38
118 O.s. miRNA families
7. Identify conserved complementarity to mRNAs
278 20mers (AtSet6)
100 A.t. miRNA loci 41/48
24 A.t. miRNA families
349 20mers (OsSet6)
193 O.s. miRNA loci 36/38
24 O.s. miRNA families
8. Verify expression of miRNAs and regulation of targets
81 A.t. miRNA loci 41/48
18 A.t. miRNA families
111 O.s. miRNA loci 36/38
18 O.s. miRNA families
9. Identify paralogs
87 A.t. miRNA loci 64/69
18 A.t. miRNA families
122 O.s. miRNA loci 83/83
18 O.s. miRNA families


B
26 nt loop
18 nt loop
14 nt loop
9 nt loop
5' 3'
MIR393a
Chr 2
Arabidopsis
MIR393b
Chr 3
Arabidopsis
MIR393
Contig 4493
Oryza


C
23 nt loop
5' 3'
MIR394a
Chr 1
Arabidopsis
MIR394b
Chr 1
Arabidopsis
MIR394
Contig 15318
Oryza

surface of leaves, indicating that the miR165/miR166 family has a conserved role in determining leaf polarity despite the morphological differences between *Arabidopsis* and maize leaves (Juarez et al., 2004). Transgenic plants with silent mutations in the miR-JAW complementary sites of TCP transcription factors arrest as seedlings with fused cotyledons and lack shoot apical meristems, while those with mutations in the miR159 complementary site of *MYB33* have upwardly curled leaves (Palatnik et al., 2003). Plants deficient in miR172-mediated regulation of *APETALA2* have altered patterns of floral organ development (Chen, 2004). Plants deficient in miR164-mediated regulation of *CUP-SHAPED COTYLEDON1* have altered patterns of embryonic, vegetative, and floral development (Mallory et al., 2004). Finally, silent mutations in the miR168 complementary site of *ARGONAUTE1* lead to misregulation of miRNA targets and numerous developmental defects (Vaucheret et al., 2004).

To gain a more complete understanding of plant miRNAs and their regulatory targets, we devised a computational procedure to identify conserved miRNA genes that were missed in previous cloning efforts, and we refined our computational method for identifying mRNA targets to increase its sensitivity. Using criteria that retain all 11 of the previously identified miRNA gene families conserved between *Arabidopsis thaliana* and *Oryza sativa*, we found 13 additional families of candidates. Molecular evidence showed that at least seven of these newly identified families of candidate miRNAs are authentic, and that at least six out of the seven mediate the cleavage of their predicted mRNA targets. These seven newly identified families are represented by 23 loci. When these are added to those identified by cloning, we count 92 miRNA loci in the *Arabidopsis* genome. Our updated analysis of the plant miRNA targets indicates a continued very strong overall bias toward transcription factors and genes involved in development. Some targets of the newly identified miRNAs, such as F-box proteins and GRL transcription factors, represent genes with demonstrated or probable roles in controlling developmental processes. Nonetheless, other newly identified miRNA targets, such as ATP sulfurylases, laccases, and superoxide dismutases, show that the range of functionalities regulated by miRNAs is broader than previously known. Furthermore, the expression of miR395, which targets genes involved in sulfate assimilation, is responsive to the sulfate concentration of the growth media, demonstrating that miRNA expression can be modulated by levels of external metabolites.

Results

Identification of 20mers in Conserved miRNA-like Hairpins

Our computational approach to identify plant miRNAs was based upon six characteristics that describe previously known plant miRNAs. (1) The base pairing of the mature miRNA to its miRNA* within the hairpin precursors is relatively consistent. In contrast, both the size of the foldback and the extent of base pairing outside of the immediate vicinity of the miRNA are highly variable among the hairpins of plant miRNAs, even among those of miRNAs from the same gene family. (2) The majority of known *Arabidopsis* miRNAs have identifiable homologs in the *Oryza sativa* genome, in which the predicted mature *Oryza* miRNAs have 0-2 base substitutions relative to their *Arabidopsis* homologs. (3) The secondary structures of known miRNA hairpins are robustly predicted by RNAfold if given a sequence sufficiently long to contain both the miRNA and the miRNA*. (4) The sequences of the *Arabidopsis* and *Oryza* hairpins are generally more conserved in the miRNA and miRNA* than in the segment joining the miRNA and miRNA*. (5) All matches to known miRNAs in the *Arabidopsis* genome, with the exception of those antisense to coding regions, have potential miRNA-like hairpins and are thus annotated as miRNA genes. (6) Most known *Arabidopsis* miRNAs are highly complementary to target mRNAs, and this complementarity is conserved to *Oryza*.

As the first step to identifying miRNAs in the genomes of *Arabidopsis thaliana* and *Oryza sativa*, we considered only those genomic portions contained in imperfect inverted repeats as defined by EINVERTED (Figure 1A, step 1). Within these 133,864 *Arabidopsis* and 410,167 *Oryza* inverted repeats were 73 of 86 reference set loci corresponding to the 24 previously reported miRNAs (refset1, Supplemental Table S1 at http://www.molecule.org/cgi/content/full/14/6/787/DC1). Secondary structures for the inverted repeats were predicted with RNAfold, and all 20mers within the inverted repeats were checked against MIRcheck, an algorithm written to identify 20mers with the potential to encode plant miRNAs (Figure 1A, step 2). MIRcheck takes as input (a) the sequence of a putative miRNA hairpin, (b) a secondary structure of the putative hairpin, and (c) a 20mer sequence within the hairpin to be considered as a potential plant miRNA. MIRcheck parameters were adjusted to restrict the total number of unpaired nucleotides (no more than 4 in the putative miRNA), the number of bulged or asymmetrically unpaired nucleotides (no more than 1 in the putative miRNA), the number of consecutive

Figure 1. Prediction of Conserved Plant miRNAs

(A) Outline of the computational approach used to identify conserved plant miRNAs. See text for description. In steps 1–8, the sensitivity is reported (blue) as the fraction of miRNA loci retained with perfect matches to previously identified miRNAs (refset1). In step 9, this fraction extends to imperfect matches to previously identified miRNAs. In the later steps, the total numbers of predicted miRNA loci are also reported (red).

(B and C) Predicted hairpin secondary structures of two newly identified miRNA families, 393 (B) and 394 (C), that target mRNAs of F-box proteins. Nucleotides in red comprise the sequence of the most common mature miRNA as deduced from PCR validation and Northern hybridization. Nucleotides in blue indicate additional portions of the hairpins predicted to have miRNA-encoding potential after identification of conserved 20mers in miRNA-like hairpins (Figure 1A, step 6), but before identification of conserved complementarity to mRNAs or experimental evaluation.

unpaired nucleotides (no more than 2 in the putative miRNA), and the length of the hairpin (at least 60 nucleotides inclusive of the putative miRNA and miRNA*). In contrast to the algorithms designed to identify metazoan miRNAs, MIRcheck has no requirements pertaining to the pattern or extent of base pairing in other parts of the predicted secondary structure. Even though these parameters were chosen to be relatively stringent, only 7 of the 73 remaining *Arabidopsis* and *Oryza* refset1 loci were lost at this step.

After removal of 20mers that overlap with repetitive elements, or which have highly biased sequence compositions, 389,648 *Arabidopsis* 20mers (*At*Set1) and 1,721,759 *Oryza* 20mers (*Os*Set1) had at least 1 locus that passed MIRcheck. We used Patscan to identify 20mers in *At*Set1 that matched at least one 20mer in *Os*Set1 with 0-2 base substitutions, considering only 20mers on the same arm of their putative hairpins (Figure 1A, step 3). 3,851 *Arabidopsis* 20mers had at least 1 *Oryza* match (*At*Set2), and 5,438 *Oryza* 20mers were matched at least once (*Os*Set2).

For the previously known plant miRNAs, RNAfold predicts a secondary structure in which the miRNA is paired to the miRNA*, provided that the flanking sequence is sufficiently long to contain the miRNA*. The presence of additional flanking sequence does not interfere with the prediction of a miRNA-like secondary structure. This robustly predicted folding is observed for all of the loci of each cloned miRNA, even though they have widely divergent flanking sequences. While recognizing that the predicted folds are unlikely to be correct in all their details, it is reasonable to propose that the overall robustness of the predicted folding might reflect an evolutionary optimization for defined folding in the plant. To eliminate candidates that do not fold as robustly as the previously known miRNAs, we required *At*Set2 and *Os*Set2 20mers to pass MIRcheck a second time after being computationally folded in the context of sequences flanking the hairpin. Patscan was used to find all matches of *At*Set2 and *Os*Set2 to their respective genomes, RNAfold was used to predict the secondary structure of each match in the context of a 500 nt genomic sequence centered on the 20mer, and each match was evaluated by MIRcheck (Figure 1A, step 4). 2,588 *Arabidopsis* 20mers (*At*Set3) and 3,083 *Oryza* 20mers (*Os*Set3) had at least one locus that passed MIRcheck. Because EINVERTED misses some hairpins and because this second MIRcheck evaluation used more relaxed cutoffs (up to 6 unpaired nt each in the putative miRNA and miRNA*), this step also recovered paralogs that were missed in steps 1 or 2.

The genomic matches to known *Arabidopsis* miRNAs are all either in hairpins or antisense to coding regions. To ensure that computationally identified miRNAs met this criterion, *Arabidopsis* 20mers were removed from the analysis if less than 50% of intergenic matches passed MIRcheck, or if more than 50% of genomic matches overlapped with repetitive sequence elements (Figure 1A, step 5), resulting in 2,506 20mers (*At*Set4). Because gene annotation in *Oryza* is poor, we could not reliably define matches as genic or intergenic. The 2,780 *Oryza* 20mers that had at least 1 locus pass MIRcheck and had no more than 50% of genomic matches in repetitive sequence elements were included in *Os*Set4.

The next step in our analysis was to identify pairs of *Arabidopsis* and *Oryza* hairpins that have miRNA-like patterns of sequence conservation (Figure 1A, step 6). MicroRNA precursors are generally most conserved in the miRNA:mRNA* portion of the hairpin, a characteristic that has been used to help identify insect miRNA genes (Lai et al., 2003). In our procedure, we retained homologous pairs for which both the miRNA and miRNA* 20mers were more conserved than any 20mer from the loop regions. Doing pairwise comparisons of the hairpins of *At*Set4 against those of *Os*Set4 resulted in 1,145 20mers (*At*Set5) with at least 1 acceptable *Oryza* homolog.

*At*Set5 was mapped to the *Arabidopsis* genome, and overlapping 20mers were joined together to form 379 sequences with miRNA-encoding potential. A single miRNA gene could be represented by up to four of these potential miRNA sequences, representing the miRNA, the miRNA*, the antisense miRNA, and the antisense miRNA*. After accounting for multiple potential miRNAs mapping to a single locus, the 379 potential miRNAs represented 228 potential miRNA loci. These 228 loci were grouped into 118 families of potential miRNA loci based on sequence similarity as determermined by blastn. Many of these newly identified miRNA candidates had patterns of secondary structure conservation resembling those of previously known plant miRNAs (Figures 1B and 1C). For many of the miRNA loci corresponding to previously reported miRNAs, the computationally identified sequences extended 1-9 nt on either side of the cloned miRNAs, although in a few cases the actual miRNA overlapped with but extended beyond the predicted sequence.

A Refined Procedure for Predicting miRNA Targets

We previously identified mRNAs containing ungapped, antisense matches to miRNAs with 0-3 mispairs (counting G:U pairs as mispairs) as probable miRNA targets (Rhoades et al., 2002). Although the majority of validated plant miRNA targets are captured by this cutoff, there are several authentic targets that are missed. For example, miR162 has a bulged nucleotide as it basepairs to the mRNA of DCL1, and miR-JAW has 4-5 mispairs to the mRNAs of several TCP transcription factors (Palatnik et al., 2003; Xie et al., 2003). In order to more thoroughly assess the mRNA-targeting potential of both known and predicted miRNAs, we developed a more sensitive computational approach to identify target candidates. It allows for gaps and more mismatches in the mRNA:miRNA duplex but requires that the miRNA complementarity be conserved between homologous *Arabidopsis* and *Oryza* mRNAs. Each miRNA complementary site was scored, with perfect matches given a score of 0, and points were added for each G:U wobble (0.5 points), each non-G:U mismatch (1 point) and each bulged nucleotide in the miRNA or target strand (2 points). To allow the same cutoffs to be applied more evenly to miRNAs of different lengths and to avoid penalizing mismatches at the ends of longer miRNAs, those miRNAs that were longer than 20 nt were broken into overlapping 20mers, with the mRNA:miRNA pair receiving the score of the most favorable 20mer.

This scoring was tested using a set of 10 unrelated

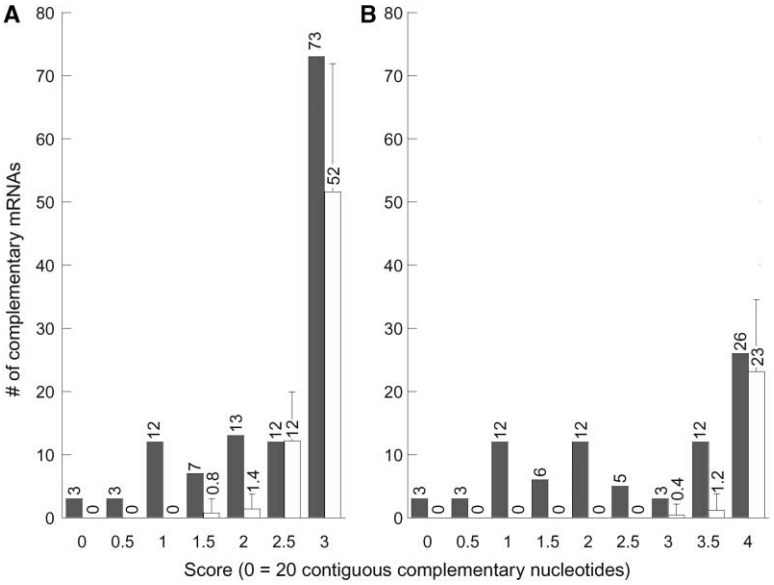


Figure 2. The Utility of Incorporating Evolutionary Conservation when Predicting Plant miRNA Targets

(A) *Arabidopsis* mRNAs with sites complementary to a set of 10 diverse miRNAs conserved between *Arabidopsis* and *Oryza* (refset2) were found and scored such that lower scores indicate fewer mismatches (see text for details). The number of mRNAs with each of the indicated scores is graphed (solid bars). Complementary sites were found and scored in the same manner for 5 cohorts of permuted miRNAs with the same dinucleotide composition as the authentic miRNAs (open bars, average number of complementary mRNAs per cohort; error bars, 2 standard deviations).

(B) mRNAs complementary to 10 miRNAs were found as in (A), with the additional requirement that at least one homologous *Oryza* mRNA be complementary to the same miRNA (solid bars). Each conserved miRNA complementary site is counted as having the either the *Arabidopsis* or *Oryza* score, whichever is higher (i.e., less complementary). Messenger RNAs with conserved complementarity to cohorts of dinucleotide shuffled miRNAs were found in the same manner (open bars, average number of complementary mRNAs; error bars, 2 standard deviations).

miRNAs that are each highly conserved (0-1 substitutions) between *Arabidopsis* and *Oryza* (refset2, Supplemental Table S1). As a control, we generated 5 cohorts of permuted miRNAs, in which each permuted miRNA has the same dinucleotide composition as the corresponding miRNA in refset2. For all 20mers from the sets of real and permuted miRNAs we searched for complementary sites in *Arabidopsis* and *Oryza* mRNAs. Compared to their shuffled cohorts, the real miRNAs had many more complementary *Arabidopsis* mRNAs with scores ≤ 2 (Figure 2A), which was in agreement with our previous results (Rhoades et al., 2002). Filtering the miRNA-complementary mRNAs to include only those conserved to *Oryza* showed that nearly all the complementary sites to authentic miRNAs with scores of ≤ 2 are conserved (Figure 2B). For the permuted miRNAs, requiring conservation reduced to nearly zero the number of complementary sites with scores of 2–3.5, whereas for the authentic miRNAs a small but significant number of sites scoring in this range were conserved (Figure 2B). Thus, adding a requirement for conservation raised the threshold at which spurious matches were found, thereby enabling confident prediction of targets that were less extensively paired to the miRNAs—in some cases forming Watson-Crick pairs to only 15 of 20 miRNA nucleotides.

Each of the conserved miRNAs had at least one predicted target with score ≤ 3.0, suggesting that the possession of predicted targets could be a criterion for screening the newly identified miRNA candidates. For each 20mer in *At*Set5 and *Os*Set5, miRNA complementary sites were found and scored (Figure 1A, step 7). As would be expected even for permuted sequences, nearly all of the *At*Set5 20mers (1,124 out of 1,145) had a complementary score of ≤ 3.0 to at least 1 *Arabidopsis* mRNA. Of these, 278 20mers (*At*Set6) had at least one homologous *Oryza* 20mer with complementarity to a homologous *Oryza* mRNA. *At*Set6 represented 24 families of potential miRNAs, which account for 100 potential miRNA loci. Eleven of these families, represented by 59 loci (including 41 refset1 loci), corresponded to all previously known miRNA families with identifiable *Oryza* homologs (Table 1), suggesting that our method also identified most of the previously unknown families that have extensive conserved complementarity in *Oryza*.

Newly Identified miRNAs Are Expressed

Our computational screen identified 13 previously unreported families of conserved miRNA candidates with conserved complementarity to mRNAs. To determine which of these putative miRNAs are expressed, we used a PCR based assay (Lim et al., 2003a, 2003b) to search for the predicted miRNAs in a library of small cDNAs (Reinhart et al., 2002). In addition to verifying the expression of the miRNAs, this assay maps the 5′ ends of the miRNAs (Table 2). Each PCR reaction used one common primer corresponding to the adaptor oligo attached to the 5′ end of all members of the library and one primer specific to the 3′ portion of the predicted miRNA. For seven miRNA families, PCR reactions resulted in products in which the specific primer was extended by at least 3 nucleotides that matched the predicted miRNA sequence. In sum, the seven newly identified miRNA families comprised 23 genomic loci in *Arabidopsis* (Table 2). All clones for families 393, 396, 397, and 398 had the same 5′ end, while for families 394, 395, and 399 miRNAs were detected with differing 5′ ends that could result from inconsistent processing of precursors transcripts from a single locus, or from differential processing of precursors from different loci. Several of these miRNA families include loci that would encode distinct but highly similar miRNAs (Table 2). Because the PCR primers overlapped with the residues that differ, it is not possible to know which variants were detected.

Six families of putative miRNAs passed all computational checks but were not validated by the PCR assay.

Table 1. Sensitivity of Computational Identification of Plant miRNA Loci

Family	*At* loci	*Os* loci
Newly identified families		
393	2/2	1/1
394	2/2	1/1
395	6/6	16/19
396	2/2	3/3
397	1/2	1/2
398	3/3	2/2
399	6/6	10/11
Previously identified conserved families		
156[a]	12/12	12/12
159/JAW[a,b,c]	3/6	7/8
160[a]	3/3	6/6
162[a,d]	2/2	2/2
164[a]	3/3	5/5
166[a]	8/9	10/12
167[a,b,d]	3/4	9/9
168[a]	2/2	1/2
169[a]	14/14	15/17
171[a,d]	4/4	7/7
172[b]	5/5	3/3
Previously identified non-conserved families		
158[a]	0/2	0
161[a]	0/1	0
163[a]	0/1	0
173[b]	0/1	0

All newly identified and previously known miRNA families are tallied. The number of loci found by de novo computational prediction (Figure 1A, through step 8) is shown (numerator) as fraction of total found by searching for near paralogs to miRNAs with verified expression (denominator). Additional details regarding the miRNA loci are reported in Tables S2 and S3 (*Arabidopsis* and *Oryza* loci, respectively). Citations for previously identified families: [a]Reinhart et al. (2002). [b]Park et al. (2002). [c]Mette et al. (2002). [d]Llave et al. (2002b).

Five of these families had a single locus in *Arabidopsis*, whereas the sixth had 14 *Arabidopsis* loci and 52 *Oryza* loci and likely represented a repetitive element not identified by RepeatMasker. Although the possibility that some of these non-validated predicted candidates are authentic cannot be ruled out, we consider it unlikely that they represent miRNA sequences.

The expression of newly identified miRNAs was also tested by Northern blot analysis. Hybridization probes were designed for representative members of the 7 miRNA families detected by the PCR assay. Probes complementary to miR393, miR394, miR396a, and miR398b detected 20-21 nt RNAs in samples from wild-type, soil-grown Columbia plants (Figure 3A), whereas probes complementary to miR395a, miR397b, and miR399b did not detect expressed small RNAs in these samples. These miRNAs that are difficult to detect on a Northern blot are likely to be expressed only at low levels or only in a subset of tissues or growth conditions.

Because miR395 is complementary to mRNAs of ATP sulfurylase (APS) proteins (Figure 5), and because the expression levels of numerous sulfate metabolizing genes are responsive to sulfate levels (Takahashi et al., 1997; Lappartient et al., 1999; Maruyama-Nakashita et al., 2003), we hypothesized that the expression of miR395 might be dependent on cellular sulfate levels. To test this, we probed RNA samples from plants grown in modified MS media containing various amount of sulfate. As seen for plants grown in soil, miR395 was not detected in the samples from plants grown in 2 mM SO_4^{2-}. However, miR395 was readily detected in the samples grown in very low sulfate (Figure 3B, 0.2 or 0.02 mM SO_4^{2-}). Induction of miR395 by low external sulfate concentrations is somewhat reminiscent of the

Table 2. Newly Identified miRNA Gene Families in *Arabidopsis*

miRNA family	miRNA gene	Chr.	Arm	miRNA Sequence
393 (PCR,N,R)	*MIR393a*	2	5′	UCCAAAGGGAUCGCAUUGAUC
	MIR393b	3	5′	UCCAAAGGGAUCGCAUUGAUC
394 (PCR,N,R)	*MIR394a*	1	5′	uUCUUUGGCAUUCUGUCCACC
	MIR394b	1	5′	uUCUUUGGCAUUCUGUCCACC
395 (PCR,N,R)	*MIR395a*	1	3′	cUGAAGUGUUUGGGGGAACUC
	MIR395b	1	3′	cUGAAGUGUUUGGGGGAACUC
	MIR395c	1	3′	cUGAAGUGUUUGGGGGAACUC
	MIR395d	1	3′	cUGAAGUGUUUGGGGGGACUC
	MIR395e	1	3′	cUGAAGUGUUUGGGGGGACUC
	MIR395f	1	3′	cUGAAGUGUUUGGGGGGACUC
396 (PCR,N,R)	*MIR396a*	2	5′	UUCCACAGCUUUCUUGAACUG
	MIR396b	5	5′	UUCCACAGCUUUCUUGAACUU
397 (PCR,R)	*MIR397a*	4	5′	UCAUUGAGUGCAGCGUUGAUG
	MIR397b	4	5′	UCAUUGAGUGCAUCGUUGAUG
398 (PCR,N,R)	*MIR398a*	2	3′	UGUGUUCUCAGGUCACCCCUU
	MIR398b	5	3′	UGUGUUCUCAGGUCACCCCUG
	MIR398c	5	3′	UGUGUUCUCAGGUCACCCCUG
399 (PCR)	*MIR399a*	1	3′	UGCCAAAGGAGAUUUGCCCUG
	MIR399b	1	3′	ccUGCCAAAGGAGAGUUGCCCUG
	MIR399c	5	3′	ccUGCCAAAGGAGAGUUGCCCUG
	MIR399d	2	3′	UGCCAAAGGAGAUUUGCCCCG
	MIR399e	2	3′	UGCCAAAGGAGAUUUGCCUCG
	MIR399f	2	3′	UGCCAAAGGAGAUUUGCCCGG

Newly identified miRNA families are listed with summary of experimental validation (PCR, PCR validation of miRNA; N, Northern blot of miRNA; R, 5′ RACE of target mRNA). The chromosome of each locus is indicated (Chr.), as is the arm of the predicted stem-loop that contains the miRNA (arm). 5′ ends of miRNAs were determined from PCR of small cDNAs, and lengths of miRNAs were inferred from mobility on Northern blots. For miRNAs not detected on Northern blots (families 397 and 399), lengths of 21 nt were assumed. For miRNA families for which multiple 5′ ends were detected by PCR, nucleotides present in some but not all clones are listed in lower case.

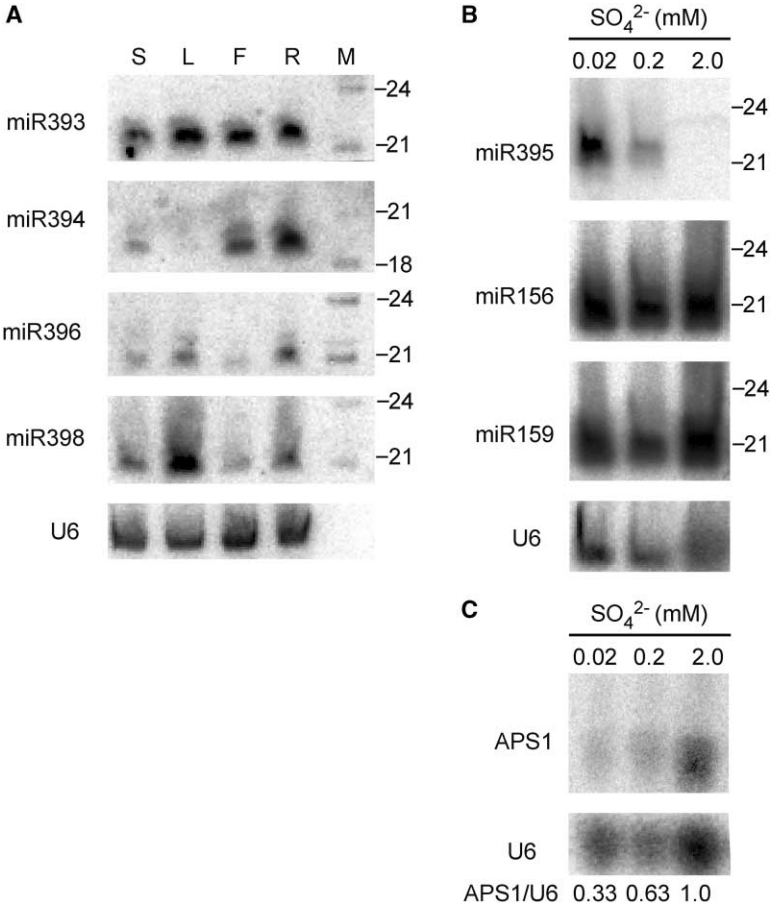


Figure 3. Expression of Newly Identified miRNAs

(A) Total RNA (30 μg) from seedlings (S), rosette leaves (L), flowers (F), and roots (R) were analyzed on a Northern blot, successively using radio-labeled DNA probes complementary to newly identified miRNAs. The lengths of 5′-phosphorylated radio-labeled RNA size markers (M) are indicated. As a loading control, the blot was probed for the U6 snRNA.

(B) miR395 is induced with low sulfate. Total RNA (40 μg) from 2-week-old Columbia plants grown on modified MS media containing the indicated concentrations of SO_4^{2-} were analyzed by Northern blot, probing for the indicated miRNAs as in (A).

(C) *APS1* mRNA decreases in low sulfate. Total RNA (10 μg) from 2-week-old plants grown on modified MS media containing the indicated concentrations of SO_4^{2-} were analyzed by Northern hybridization using randomly primed body-labeled DNA probes corresponding to exon 1 of the *APS1* mRNA. Normalized ratios of *APS1* mRNA to U6 splicosomal RNA are indicated.

starvation-associated miR-234 increase that has been observed in nematodes (Lim et al., 2003b), although the miR395 induction (greater than 100-fold) is much more striking than that of miR-234 (2-fold). We examined whether *APS1* expression changed in the conditions that induced miR395, and found that its expression decreased when miR395 increased, as would be expected if *APS1* was a cleavage target of miR395 (Figure 3C).

Experimental Verification of miRNA Targets

MicroRNAs, like small interfering RNAs (siRNAs, Elbashir et al., 2001), can direct the cleavage of their mRNA targets when these messages have extensive complementarity to the miRNAs (Hutvagner and Zamore, 2002; Llave et al., 2002b; Tang et al., 2003; Yekta et al., 2004). This miRNA-directed cleavage can be detected by using a modified form of 5′-RACE (rapid amplification of cDNA ends) because the 3′ product of the cleavage has two diagnostic properties: (1) a 5′ terminal phosphate, making it a suitable substrate for ligation to an RNA adaptor using T4 RNA ligase, and (2) a 5′ terminus that maps precisely to the nucleotide that pairs with the tenth nucleotide of the miRNA (Llave et al., 2002b; Kasschau et al., 2003). To examine whether any of the newly identified miRNAs can direct cleavage of their predicted targets in vivo, we isolated RNA from vegetative and floral tissues and performed the 5′-RACE procedure using primers specific to the predicted targets. For 19 predicted targets the 5′-RACE PCR yielded a distinct band of the predicted size on an agarose gel, which was isolated, cloned, and sequenced. In all 19 cases the most common 5′ end of the mRNA fragment mapped to the nucleotide that pairs to the tenth nucleotide of one of the miRNAs validated by PCR (Figure 4), indicating cleavage at sites precisely analogous to those seen for other miRNA targets (Llave et al., 2002b; Aukerman and Sakai, 2003; Kasschau et al., 2003; Palatnik et al., 2003; Xie et al., 2003; Vazquez et al., 2004; Mallory et al., 2004), as well as for RNAs complementary to siRNAs and metazoan miRNAs (Elbashir et al., 2001; Hutvagner and Zamore, 2002; Yekta et al., 2004). These observations also corroborate the 5′ ends of the miRNAs as mapped by PCR (Table 2).

Identification of miRNA Paralogs

Our computational approach found 81 miRNA loci from 18 miRNA families (Table 1; Supplemental Table S2). We searched for additional members of these families by searching the *Arabidopsis* genome for near matches (0-3) to the miRNAs of these 81 loci (Figure 1A, step 9). After manual inspection for potential hairpin-like secondary structures, this identified six additional loci in miRNA families that are conserved to *Oryza*. Together with the five loci in miRNA families without apparent *Oryza* homologs, this brings to 92 the total number of *Arabidopsis* loci that meet the criteria for designation

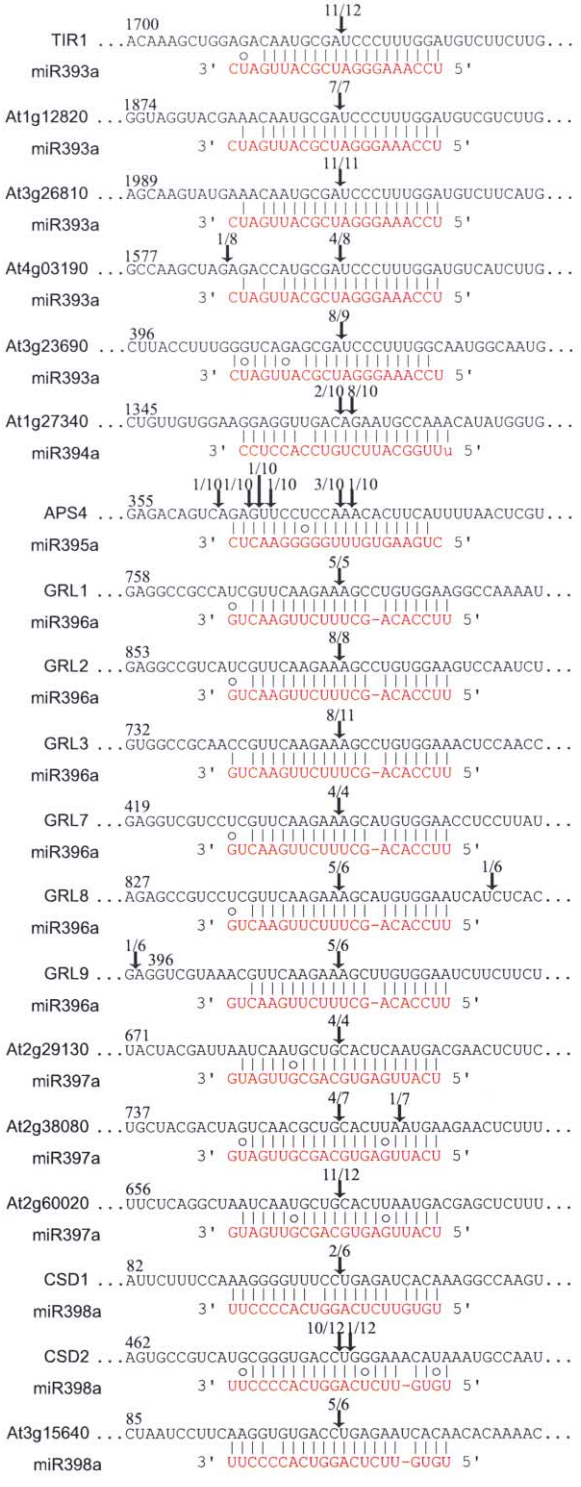


Figure 4. Experimental Verification of Predicted miRNA Targets

Each top strand (black) depicts a miRNA complementary site, and each bottom strand depicts the miRNA (red). Watson-Crick pairing (vertical dashes) and G:U wobble pairing (circles) are indicated. Arrows indicate the 5′ termini of mRNA fragments isolated from plants, as identified by cloned 5′-RACE products, with the frequency of clones shown. Only cloned sequences that matched the correct gene and had 5′ ends within a 100 nt window centered on the miRNA complmentary site are counted. The miRNA sequence shown corresponds to the most common miRNA suggested by miRNA PCR validation (Table 2). For miR394, the 5′ end of a less common variant (1 out of 4 PCR clones) is indicated in lower case and corresponds to the most commonly cloned cleavage product.

as miRNA genes (Ambros et al., 2003a) (Supplemental Table S2). As is generally the case with computational gene prediction, some of these might be pseudogenes. Our miRNA-finding algorithm found 88% of these de novo, and 93% of those with *Oryza* homologs. These *Arabidopsis* genes correspond to 122 *Oryza* miRNA genes, of which 111 (91%) were found de novo by our algorithm (Figure 1A, step 9; Supplemental Table S3).

As has been previously observed for numerous animal miRNAs (Lagos-Quintana et al., 2001; Lau et al., 2001), we find that some plant miRNA genes are clustered in the genome, most strikingly the genes of the 395 family. In *Arabidopsis*, miRNAs of the 395 family are located in two clusters, each containing three hairpins within 4 kb (Figure S1A). In each cluster, two *MIR395* hairpins are on one strand while the third is on the opposite. Thus each cluster could not be expressed as a single primary transcript, but could be expressed as two transcripts sharing common regulatory elements. The *Oryza MIR395* hairpins are also clustered, but with a different arrangement than in *Arabidopsis*. The two largest *Oryza MIR395* clusters contain seven and six hairpins, respectively, within 1 kb, with all hairpins encoded on the same strand of DNA (Figure S1B). These clusters are likely expressed as transcripts containing multiple miRNAs, an idea supported by *Oryza* EST CA764701, which contains four miR395 hairpins.

Prediction of Conserved miRNA Targets

Having refined our computational method to more sensitively predict plant miRNA targets, we applied it to the prediction of conserved mRNA targets of all known *Arabidopsis* and *Oryza* miRNAs (Figure 5). Control experiments with refset2 and 5 sets of permuted miRNAs suggested that a score cutoff of ≤3.5 was appropriate to identify conserved miRNA targets with high sensitivity and selectivity. However, when searching for targets of the entire set of miRNAs, this cutoff identified a number of mRNAs for which miRNA-mediated cleavage products could not be found by 5′-RACE. Thus, a cutoff of ≤3 was chosen to minimize the number of non-authentic targets. All previously validated miRNA targets are identified at this level of sensitivity, although several newly validated targets have scores of 3.5 in one or both species and are not retained using this cutoff. Thus, there is still a threshold at which it is difficult to distinguish authentic targets from potentially spurious complementarity without experimental verification. Nonetheless, a score of <3.0 in our refined method identifies targets with very high confidence (Figure 2).

Plant miRNAs Are Deeply Conserved

MicroRNAs conserved between the dicot *Arabidopsis thaliana* and the monocot *Oryza sativa* are likely to be found in most flowering plants. Homologs of miR-JAW and miR-JAW complemenary sites have been found in ESTs from numerous angiosperms (Palatnik et al., 2003). To look for evidence of other miRNAs in additional plant

miRNA family	Target protein class	Target genes
156	Squamosa-promoter Binding Protein (SBP)-like transcription factors	*At1g27360/SPL11*(1), *At1g27370/SPL10*(1)[a], *At1g53160/SPL4*(2), *At1g69170/SPL4*(1), *At2g33810/SPL3*(1.5), *At2g42200/SPL9*(1), *At3g15270/SPL5*(3), *At3g57920*(1), *At5g43270/SPL2*(1)[b], *At5g50570*(1), *At5g50670*(1)
159/JAW	MYB transcription factors	*At2g26950/MYB104*(1.5), *At2g26960/MYB81*(2.5), *At2g32460/MYB101*(1.5), *At3g11440/MYB65*(1.5)[a,c], At3g60460(1.5), *At4g26930/MYB97*(2.5), *At5g06100/MYB33*(1.5)[c], *At5g55020/MYB120*(2)
	TCP transcription factors	*At1g30210/TCP24*(2.5)[c], *At1g53230/TCP3*(3)[c], *At2g31070/TCP10*(2.5)[c], *At3g15030/TCP4*(2.5)[c], *At4g18390/TCP2*(2.5)[c]
160	Auxin Response Factors (ARF transcription factors)	*At1g77850/ARF17*(0.5)[a,b], *At2g28350/ARF10*(1)[b], *At4g30080/ARF16*(1.5)
162	DICER-LIKE 1	*At1g01040/DCL1*(2)[d]
164	NAC domain transcription factors	*At1g56010/NAC1*(1)[e], *At3g15170/CUC1*(1)[b,e], *At5g07680*(1.5)[e], *At5g39610*(2), *At5g53950/CUC2*(1)[a,b,e], *At5g61430*(1.5)[e]
166	HD-Zip transcription factors	*At1g30490/PHV*(1.5)[f], *At1g52150/ATHB-15*(1.5), *At2g34710/PHB*(1.5)[f], *At4g32880/ATHB-8*(1.5), *At5g60690/REV*(1.5)[g]
167	Auxin Response Factors (ARF transcription factors)	*At1g30330/ARF6*(2), *At5g37020/ARF8*(2)[a]
168	ARGONAUTE	*At1g48410/AGO*(2.5)[a,h]
169	CCAAT Binding Factor (CBF) HAP2-like transcription factors	*At1g17590*(1.5), *At1g54160*(2), *At1g72830*(1.5), *At3g05690*(1.5), *At3g20910*(2), *At5g06510*(1.5), *At5g12840*(1.5)
171	SCARECROW-like transcription factors	*At2g45160*(0), *At3g60630*(0)[a,i], *At4g00150/SCL6*(0)[i]
172	APETALA2-like trancription factors	*At2g28550/TOE1*(1.5)[b,j], *At2g39250*(1), *At4g36920/AP2*(0.5)[b,j,k], *At5g60120/TOE2*(0.5)[b,j], *At5g67180/TOE3*(1.5)[b]
393	F-box proteins	*At1g12820*(1), *At3g26810*(1), *At3g62980/TIR1*(1.5), *At4g03190*(2.5)
	bHLH transcription factor	*At3g23690*(2)*
394	F-box protein	*At1g27340*(1)
395	ATP sulfurylases	*At3g22890/APS1*(1.5), *At4g14680/APS3*(1.5), *At5g43780/APS4*(0.5)
396	Growth Regulating Factor (GRL) transcription factors	*At2g22840/GRL1*(3), *At2g36400/GRL3*(3), *At2g45480/GRL9*(3), *At3g52910/GRL4*(3), *At4g24150/GRL8*(3), *At4g37740/GRL2*(3), *At5g53660/GRL7*(3)
	Rhodenase-like protein	*At2g40760*(2.5)
	Kinesin-like protein B	*At4g27180/ATK2*(3)
397	Laccases	*At2g29130*(0.5), *At2g38080*(1), *At5g60020*(1)
	Beta-6 tubulin	*At5g12250*(3)
398	Copper superoxide dismutases	*At1g08830/CSD1*(3),*At2g28190/CSD2*(3.5)*
	Cytochrome C oxidase subunit V	*At3g15640*(3)*
399	Phosphate transporter	*At3g54700*(2)

Figure 5. Converved Predicted miRNA Targets

All predicted miRNA targets with scores of 3.0 or less in both *Arabadopsis* and *Oryza* are listed. For each target gene the score of the best scoring 20mer from any member of the miRNA family is given in parentheses. Predicted targets with scores greater than 3.0 in either *Arabadopsis* or *Oryza* but that have been validated by 5′-RACE are also listed and marked with an asterisk. Genes in red were validated as miRNA targets by 5′-RACE experiments in this work. Genes in blue are validated as miRNA targets by previous work. Additional information on target genes can be found at www.arabadopsis.org. a, Vazquez et al., 2004; b, Kasschau et al., 2003; c, Palatnik et al., 2003; d, Xie et al., 2003; e, Mallory et al., 2004; f, Tang et al., 2003; g, Emery et al., 2003; h, Vaucheret et al., 2004; i, Llave et al., 2002a; j, Aukerman and Sakai, 2003; k, Chen, 2003.

species, we searched for ESTs representing potential homologs of *Arabidopsis* and *Oryza* miRNAs, defined here as having 19/20 nt matches and a predicted fold-back that passes MIRcheck. This search identified 187 putative miRNA homologs in the ESTs (Supplemental Table S4). A large majority of these appear to be authentic, in that the 10 miRNAs in refset2 each had on average 9.7 EST matches that passed MIRcheck, whereas the set of 50 permuted miRNAs averaged only 0.04 matches that passed MIRcheck. For all 18 miRNA families that are conserved between *Arabidopsis* and *Oryza,* potential miRNA precursors were found in at least one additional angiosperm species (Supplemental Table S4). For miRNAs that are not conserved between *Arabidopsis* and *Oryza*, no homologous miRNAs in additional species were identified, suggesting that the lack of conservation in *Oryza* is a consequence of recent emergence rather than loss in the *Oryza* lineage. We also searched for matches to experimentally confirmed miRNA complementary sites in ESTs encoding proteins homologous to *Arabidopsis*

targets (blastx score $<10^{-6}$). For all conserved miRNA families with validated miRNA targets, conserved miRNA complementary sites (19/20 nt matches) were found in at least one additional angiosperm (Supplemental Table S5). On average, the miRNA complementary sites from 17 representative *Arabidopsis* miRNA targets were each conserved in 191 homologous ESTs, representing 14 species. This is far more than would be expected by chance; when repeating the analysis using 170 sites chosen at random from the same Arabidopsis mRNAs, the average number of ESTs and species were 2.6 and 0.5, respectively.

MicroRNAs of the 166 family, as well as their binding sites in mRNAs of HD-ZIP proteins, predate the emergence of seed plants (Floyd and Bowman, 2004). We found nine miRNA families (156, 160, 166, 167, 393, 395, 396, 397, and 398) that had complementary sites conserved in gymnosperms, while a miR171 complementary site was conserved in a *SCL* mRNA from a fern (*Ceratopteris richarii*). In addition, a potential miRNA hairpin of the 159/JAW family was present in an EST from moss (*Physomitrella patens*). These data suggest that multiple miRNAs have deep origins in plant phylogeny.

Discussion

The Scope of miRNAs Conserved between Dicots and Monocots

A combination of computational prediction and experimental verification identified seven families of sequences that had not previously been identified as miRNAs. A set of 2088 small RNAs from *Arabidopsis* was recently reported (Xie et al., 2004) (http://gac.bcc.orst.edu/smallRNA/). Sequences corresponding to miR397a, miR398b, and miR399b were contained in this dataset, each having been cloned a single time, although none were annotated as miRNAs. The cloning of miR397 and miR399, which were not detected by Northern blot, corroborates their expression as determined by PCR.

Families 393, 394, 395, and 396 are absent from the reported sets of cloned, sequenced small RNAs. These are each detectable by Northern analysis, and as with families 397, 398, and 399 were detected by PCR in our library of small cDNAs used for cloning. Therefore they would have been found eventually by sequencing enough small cDNAs. However, given that other miRNAs have been cloned hundreds of times (Xie et al., 2004), it seems that all seven newly identified miRNA families are relatively rare in the tissues and growth conditions from which small RNAs have been cloned. They may represent miRNAs that are needed at low levels, or whose expression is limited to rare cell types or particular growth conditions. The expression of miR395 is greatly increased by sulfate starvation; other miRNAs with seemingly low expression may also be inducible by metabolite levels or environmental stimuli. It is the identification of these difficult to clone but potentially important miRNAs that makes computational prediction a useful complement to cloning of small RNAs.

The sensitivity of our computational approach, which found all 11 conserved miRNA families previously identified through cloning, suggests that most plant miRNAs with properties similar to previously cloned miRNAs have been identified. MicroRNA genes not found by our analysis are likely to fall into several categories. One set will be those without apparent conservation to *Oryza*. This describes four families of currently known *Arabidopsis* miRNAs (158, 161, 163, and 173). It is difficult to estimate how many additional non-conserved miRNA families exist in either species, but the observation that most of the cloned plant miRNAs have readily identified *Oryza* homologs indicates either that there are no more than a handful of non-conserved miRNAs remaining to be identified or that non-conserved miRNAs are disproportionally poorly expressed in plants.

Another set of false negatives will be miRNA families that are conserved between *Arabidopsis* and *Oryza* but were missed by our analysis. Most steps in our analysis have the potential to lose authentic miRNA genes. The parameters and cutoffs we used were chosen to be slightly more relaxed than what was needed to retain most loci corresponding to the 11 previously known miRNAs families with *Oryza* homologs in refset1. They found at least one member of each family and 92% (59/64) of all loci in these families. A similar percentage of loci, 96%, were correctly identified for newly discovered miRNA families (22/23), suggesting that our parameters are not over fitted. Relaxing the parameters of MIRcheck (Figure 1, steps 2 and 4) to allow up to two asymmetric bulges, shorter hairpins (as short as 54 nt), and an additional mismatch did not identify any additional verifiable miRNAs (data not shown). Nonetheless, the low number of previously identified *Arabidopsis* miRNA gene families (15) precluded splitting the miRNAs into a training set and test set, as was done in our metazoan analysis to evaluate the degree of overtraining and enable firm estimates of the number of genes remaining to be identified (Lim et al., 2003b). MicroRNA families with few members would be more prone to being missed. For example, *MIR393* and *MIR394* each have only one identified locus in *Oryza*; either would have been missed if their *Oryza* locus had been among the fraction of authentic miRNA loci not identified as an inverted repeat or that did not pass MIRcheck, whereas miRNAs that were members of larger gene families that have multiple *Oryza* homologs were identified even though some *Oryza* homologs were missed. The observation that some miRNA primary transcripts are spliced (Aukerman and Sakai, 2003) raises the possibility that some miRNA transcripts might have an intron within the hairpin precursor, which could prevent their identification in our analysis of genomic DNA. Furthermore, any unknown miRNA family that systematically had a pattern of base pairing that failed MIRcheck would also have been lost, but there is no reason to suspect that this was a widespread problem.

More significant uncertainty in plant miRNA gene number arises from the 94 families of candidate miRNAs that had conserved miRNA-like hairpins but lacked extensive and conserved complementarity to mRNAs. Some of these candidates may be authentic miRNAs with different modes of target recognition. For example, any plant miRNA that recognizes all its target mRNAs in a manner similar to that of most animal miRNAs, that is, by recognizing its targets predominantly through "seed matches" (Lewis et al., 2003), would have been missed. However, these 94 families of candidates generally differ from the validated miRNAs in that only seven

of the 94 have potential orthologs recognizable within the EST database (data not shown), whereas all 18 validated families have such orthologs. Nonetheless, further analysis will be required before a meaningful upper bound on the number of plant miRNA genes can be estimated. The 92 loci tabulated to date, when considered together with the assumption that a few others might remain undetected because they are refractory to both cloning and computation, places a lower bound on the number of *Arabidopsis* miRNA genes at ~100, or ~0.4% of the predicted *Arabidopsis* genes—a percentage somewhat lower than that of animals. The plant miRNAs are generally in larger, more highly related families, further reducing the relative complexity of known miRNA sequences when compared to those of animals. Of course, when considering the vast number of distinct ~22 nt RNAs that have been cloned from plants, which might be endogenous siRNAs but are not miRNAs, the diversity of small silencing RNAs in plants could exceed that in animals.

The Targets of Newly Identified miRNAs

The detection of the RNA fragments diagnostic of miRNA-directed cleavage confirms in planta these 19 newly identified miRNA-target interactions. However, these 5′-RACE results do not rule out the possibility that the predominant mode of silencing is translational inhibition. 5′-RACE experiments demonstrate that miR172 directs the cleavage of some *APETALA2* mRNA molecules, even though the predominate mode of repression appears to be translational inhibition (Aukerman and Sakai, 2003; Kasschau et al., 2003; Chen, 2004). Nonetheless, for all the other plant miRNA targets examined, inhibition of the miRNA pathway leads to increased accumulation of target mRNA (Kasschau et al., 2003; Vaucheret et al., 2004; Vazquez et al., 2004), suggesting that mRNA cleavage typically plays a significant regulatory role, although in these cases augmentation by translational repression cannot be ruled out. The same is likely to be true for our newly identified targets.

Some of the newly identified targets resemble those of previous predictions with regard to their proven or inferred roles in regulating developmental processes (Figure 5). miR396 targets seven Growth Regulating Factor genes, which are putative transcription factors that regulate cell expansion in leaf and cotyledon (Kim et al., 2003). miR393 and miR394 both target the messages of F-box proteins, which in turn target specific proteins for proteolysis by making them substrates for ubiquitination by SCF E3 ubiquitin ligases (Vierstra, 2003). *At2g27340*, targeted by miR394, is in the same subfamily of F-box genes as UNUSUAL FLORAL ORGANS (*UFO*) (Gagne et al., 2002), which is involved in floral initiation and development (Wilkinson and Haughn, 1995; Samach et al., 1999). miR393 targets four closely related F-box genes, including TRANSPORT INHIBITOR RESPONSE1 (*TIR1*), which targets AUX/IAA proteins for proteolysis in an auxin-dependent manner and is necessary for auxin-induced growth processes (Ruegger et al., 1998; Gray et al., 2001). These five F-box genes constitute a newly identified biochemical class of miRNA targets.

The identification of *TIR1* as a miRNA target implies that miRNAs regulate auxin responsiveness at multiple points. Other auxin-related miRNA targets include Auxin Response Factors (miR160 and miR167) (Rhoades et al., 2002; Kasschau et al., 2003), which are thought to regulate transcription in response to auxin (Ulmasov et al., 1999), and *NAC1* (miR164) (Rhoades et al., 2002; Mallory et al., 2004), which promotes auxin-induced lateral root growth downstream of *TIR1* (Xie et al., 2000). Finally, in addition to targeting F-box genes, miR393 also targets At3g23690, a basic helix-loop-helix transcription factor with homology to GBOF-1 from tulip, which Genbank annotates as auxin inducible.

Other newly identified miRNA targets have less obvious connections to the control of developmental patterning (Figure 5). miR397 targets putative laccases, members of a family of enzymes with numerous described roles in fungal biology but without well-defined roles in plant biology (Mayer and Staples, 2002). miR399 targets two copper superoxide dismutases, CSD1 and CSD2, enzymes that protect the cell against radicals and whose expression patterns respond to oxidative stress (Kliebenstein et al., 1998).

The most definitive example of a plant miRNA operating outside the gene regulatory circuitry controlling development is miR395. miR395 targets the ATP sulfurylases, APS1, APS3, and APS4, enzymes that catalyze the first step of inorganic sulfate assimilation (Leustek, 2002). The observations that the expression of miR395 depends on sulfate concentration and that *APS1* expression declines with increasing miR395 corroborate the idea that this miRNA regulates sulfate metabolism (Figure 3).

Our systematic analysis, which probably has identified most plant miRNAs with conserved and extensive complementarity to plant messages, including those that are expressed at very low levels during lab growth conditions, allows us to revisit the question of what this class of tiny regulatory RNAs is generally doing in plants. As before (Rhoades et al., 2002), we find an overwhelming propensity for targeting messages of known or suspected plant transcription factors (63 of 83, or 76% of genes in Figure 5) and similar propensity for targeting messages of genes with known or suspected roles in plant development (70 of 83, or 84% of genes in Figure 5). A strong propensity to target developmental regulators differs from what has been seen in mammals (Lewis et al., 2003). Nonetheless, the conserved targets of plant miRNAs extend beyond the regulatory circuitry of development. The discovery that miRNAs regulate genes such as ATP sulfurylases, laccases, and superoxide dismutases shows that miRNAs also have an ancient role in regulating other aspects of plant biology.

Experimental Procedures

See Supplemental Material at http://www.molecule.org/cgi/content/full/14/6/787/DC1 for details of the computational miRNA prediction method and sequences of primers used.

PCR Validation of miRNAs

We used a PCR based assay to detect expression and map the 5′ ends of predicted miRNAs (Lim et al., 2003b). miRNAs were PCR amplified out of a library of small cDNAs from leaf, flower, and seedling flanked by 5′ and 3′ adaptor oligos (Reinhart et al., 2002). Each PCR reaction used one common primer corresponding the 5′

adaptor oligo and one specific primer antisense to the 3′ portion of the predicted miRNA.

RNA Purification and Northern Hybridization

RNA was isolated as previously described (Vance, 1991). For developmental Northerns, 30 μg per lane of total RNA from soil grown Colombia plants was separated by 15% polyacrylamide electrophoresis and blotted to a nylon membrane.

For plants grown on media, Columbia plants were grown in long-day conditions on modified MS/agarose media, containing 0.8% Agarose-LE (USBiochem), in which the SO_4^{2-} containing salts of minimal MS media were replaced with their chloride counterparts and the media supplemented with 20 μM to 2 mM $2(NH_4)SO_4$. RNA was harvested from 2-week-old plants. For miRNA Northerns, 40 μg per lane was used in Northern blots as above. For miR393, miR394, miR396a, and miR398b, end-labeled antisense DNA probes were used. For miR395a, miR397b, and miR399b, higher specific activity Starfire (Integrated DNA technologies) probes were used. MicroRNA Northerns were hybridized and washed as previously described (Lau et al., 2001). For mRNA Northerns, 10 μg per lane was separated by agarose electrophoresis and blotted as described (Mallory et al., 2001). Probes to exon 1 of APS1 were made using the Megaprime DNA labeling system (Amersham).

5′-RACE Analysis

5′-RACE was performed on poly(A)-selected RNA from Columbia inflorescences and rosette leaves using the GeneRacer Kit (Invitrogen) as described (Kasschau et al., 2003), except that nested PCR was done for each gene, with each round of PCR using one gene-specific primer and the GeneRacer 5′ Nested Primer. For each gene we designed gene-specific primers that were 180–450 bp away from the predicted miRNA binding site. PCR reactions were separated by agarose gel electrophoresis, and distinct bands of the appropriate size for miRNA-mediated cleavage were purified (excised gel slices corresponded to a size range of ~100 basepairs), cloned, and sequenced.

Acknowledgments

We thank Mike Axtell for the 5′-RACE library, Ramya Rajogopalan for the library of 18-28 nt cDNAs and Allison Mallory and other Bartel lab members for helpful discussions. This work was supported by grants from the NIH.

Received: April 9, 2004
Revised: May 24, 2004
Accepted: May 28, 2004
Published: June 18, 2004

References

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. (2003a). A uniform system for microRNA annotation. RNA *9*, 277–279.

Ambros, V., Lee, R.C., Lavanway, A., Williams, P.T., and Jewell, D. (2003b). MicroRNAs and other tiny endogenous RNAs in *C. elegans*. Curr. Biol. *13*, 807–818.

Aukerman, M.J., and Sakai, H. (2003b). Regulation of flowering time and floral organ identity by a MicroRNA and its *APETALA2*-like target genes. Plant Cell *15*, 2730–2741.

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

Bartel, B., and Bartel, D.P. (2003). MicroRNAs: at the root of plant development? Plant Physiol. *132*, 709–717.

Chen, X. (2004). A MicroRNA as a translational repressor of *APETALA2* in *Arabidopsis* flower development. Science *303*, 2022–2025.

Elbashir, S.M., Lendeckel, W., and Tuschl, T. (2001). RNA interference is mediated by 21- and 22-nucleotide RNAs. Genes Dev. *15*, 188–200.

Emery, J.F., Floyd, S.K., Alvarez, J., Eshed, Y., Hawker, N.P., Izhaki, A., Baum, S.F., and Bowman, J.L. (2003). Radial patterning of *Arabidopsis* shoots by class III HD-ZIP and *KANADI* genes. Curr. Biol. *13*, 1768–1774.

Enright, A.J., John, B., Gaul, U., Tuschl, T., Sander, C., and Marks, D.S. (2003). MicroRNA targets in *Drosophila*. Genome Biol. *5*, R1.

Floyd, S.K., and Bowman, J.L. (2004). Gene regulation: ancient microRNA target sequences in plants. Nature *428*, 485–486.

Gagne, J.M., Downes, B.P., Shiu, S.H., Durski, A.M., and Vierstra, R.D. (2002). The F-box subunit of the SCF E3 complex is encoded by a diverse superfamily of genes in *Arabidopsis*. Proc. Natl. Acad. Sci. USA *99*, 11519–11524.

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.

Gray, W.M., Kepinski, S., Rouse, D., Leyser, O., and Estelle, M. (2001). Auxin regulates SCF(TIR1)-dependent degradation of AUX/IAA proteins. Nature *414*, 271–276.

Grishok, A., Pasquinelli, A.E., Conte, D., Li, N., Parrish, S., Ha, I., Baillie, D.L., Fire, A., Ruvkun, G., and Mello, C.C. (2001). Genes and mechanisms related to RNA interference regulate expression of the small temporal RNAs that control *C. elegans* developmental timing. Cell *106*, 23–34.

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.

Hutvagner, G., and Zamore, P.D. (2002). A microRNA in a multiple-turnover RNAi enzyme complex. Science *297*, 2056–2060.

Hutvagner, G., McLachlan, J., Pasquinelli, A.E., Balint, E., Tuschl, T., and Zamore, P.D. (2001). A cellular function for the RNA-interference enzyme Dicer in the maturation of the let-7 small temporal RNA. Science *293*, 834–838.

Juarez, M.T., Kui, J.S., Thomas, J., Heller, B.A., and Timmermans, M.C. (2004). microRNA-mediated repression of *rolled leaf1* specifies maize leaf polarity. Nature *428*, 84–88.

Kasschau, K.D., Xie, Z., Allen, E., Llave, C., Chapman, E.J., Krizan, K.A., and Carrington, J.C. (2003). P1/HC-Pro, a viral suppressor of RNA silencing, interferes with *Arabidopsis* development and miRNA unction. Dev. Cell *4*, 205–217.

Ketting, R.F., Fischer, S.E., Bernstein, E., Sijen, T., Hannon, G.J., and Plasterk, R.H. (2001). Dicer functions in RNA interference and in synthesis of small RNA involved in developmental timing in *C. elegans*. Genes Dev. *15*, 2654–2659.

Khvorova, A., Reynolds, A., and Jayasena, S.D. (2003). Functional siRNAs and miRNAs exhibit strand bias. Cell *115*, 209–216.

Kidner, C.A., and Martienssen, R.A. (2004). Spatially restricted microRNA direct leaf polarity through *ARGONAUTE1*. Nature *428*, 81–84.

Kim, J.H., Choi, D., and Kende, H. (2003). The AtGRF family of putative transcription factors is involved in leaf and cotyledon growth in *Arabidopsis*. Plant J. *36*, 94–104.

Kliebenstein, D.J., Monde, R.A., and Last, R.L. (1998). Superoxide dismutase in *Arabidopsis*: an eclectic enzyme family with disparate regulation and protein localization. Plant Physiol. *118*, 637–650.

Lagos-Quintana, M., Rauhut, R., Lendeckel, W., and Tuschl, T. (2001). Identification of novel genes coding for small expressed RNAs. Science *294*, 853–858.

Lai, E.C. (2003). microRNAs: runts of the genome assert themselves. Curr. Biol. *13*, R925–R936.

Lai, E.C., Tomancak, P., Williams, R.W., and Rubin, G.M. (2003). Computational identification of *Drosophila* microRNA genes. Genome Biol. *4*, R42.

Lappartient, A.G., Vidmar, J.J., Leustek, T., Glass, A.D., and Touraine, B. (1999). Inter-organ signaling in plants: regulation of ATP sulfurylase and sulfate transporter genes expression in roots mediated by phloem-translocated compound. Plant J. *18*, 89–95.

Lau, N.C., Lim, L.P., Weinstein, E.G., and Bartel, D.P. (2001). An abundant class of tiny RNAs with probable regulatory roles in *Caenorhabditis elegans*. Science *294*, 858–862.

Lee, R.C., and Ambros, V. (2001). An extensive class of small RNAs in *Caenorhabditis elegans*. Science *294*, 862–864.

Lee, Y., Jeon, K., Lee, J.T., Kim, S., and Kim, V.N. (2002). MicroRNA maturation: stepwise processing and subcellular localization. EMBO J. *21*, 4663–4670.

Lee, Y., Ahn, C., Han, J., Choi, H., Kim, J., Yim, J., Lee, J., Provost, P., Radmark, O., Kim, S., and Kim, V.N. (2003). The nuclear RNase III Drosha initiates microRNA processing. Nature *425*, 415–419.

Leustek, T. (2002). Sulfate Metabolism. In The *Arabadopsis* Book, C.R. Somerville and E.M. Meyerowitz, eds. (Rockville, MD: America Society of Plant Biologists).

Lewis, B.P., Shih, I.H., Jones-Rhoades, M.W., Bartel, D.P., and Burge, C.B. (2003). Prediction of mammalian microRNA targets. Cell *115*, 787–798.

Lim, L.P., Glasner, M.E., Yekta, S., Burge, C.B., and Bartel, D.P. (2003a). Vertebrate microRNA genes. Science *299*, 1540.

Lim, L.P., Lau, N.C., Weinstein, E.G., Abdelhakim, A., Yekta, S., Rhoades, M.W., Burge, C.B., and Bartel, D.P. (2003b). The microRNAs of *Caenorhabditis elegans*. Genes Dev. *17*, 991–1008.

Llave, C., Kasschau, K.D., Rector, M.A., and Carrington, J.C. (2002a). Endogenous and silencing-associated small RNAs in plants. Plant Cell *14*, 1605–1619.

Llave, C., Xie, Z., Kasschau, K.D., and Carrington, J.C. (2002b). Cleavage of Scarecrow-like mRNA targets directed by a class of *Arabidopsis* miRNA. Science *297*, 2053–2056.

Mallory, A.C., Ely, L., Smith, T.H., Marathe, R., Anandalakshmi, R., Fagard, M., Vaucheret, H., Pruss, G., Bowman, L., and Vance, V.B. (2001). HC-Pro suppression of transgene silencing eliminates the small RNAs but not transgene methylation or the mobile signal. Plant Cell *13*, 571–583.

Mallory, A.C., Dugas, D.V., Bartel, D.P., and Bartel, B. (2004). MicroRNA regulation of NAC-domain targets is required for proper formation and separation of adjacent embyonic, vegatative, and floral organs. Curr. Biol. *14*, in press.

Maruyama-Nakashita, A., Inoue, E., Watanabe-Takahashi, A., Yamaya, T., and Takahashi, H. (2003). Transcriptome profiling of sulfur-responsive genes in *Arabidopsis* reveals global effects of sulfur nutrition on multiple metabolic pathways. Plant Physiol. *132*, 597–605.

Mayer, A.M., and Staples, R.C. (2002). Laccase: new functions for an old enzyme. Phytochemistry *60*, 551–565.

McConnell, J.R., Emery, J., Eshed, Y., Bao, N., Bowman, J., and Barton, M.K. (2001). Role of *PHABULOSA* and *PHAVOLUTA* in determining radial patterning in shoots. Nature *411*, 709–713.

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.

Mourelatos, Z., Dostie, J., Paushkin, S., Sharma, A., Charroux, B., Abel, L., Rappsilber, J., Mann, M., and Dreyfuss, G. (2002). miRNPs: a novel class of ribonucleoproteins containing numerous microRNAs. Genes Dev. *16*, 720–728.

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.

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.

Reinhart, B.J., Weinstein, E.G., Rhoades, M.W., Bartel, B., and Bartel, D.P. (2002). MicroRNAs in plants. Genes Dev. *16*, 1616–1626.

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

Ruegger, M., Dewey, E., Gray, W.M., Hobbie, L., Turner, J., and Estelle, M. (1998). The TIR1 protein of *Arabidopsis* functions in auxin response and is related to human SKP2 and yeast grr1p. Genes Dev. *12*, 198–207.

Samach, A., Klenz, J.E., Kohalmi, S.E., Risseeuw, E., Haughn, G.W., and Crosby, W.L. (1999). The *UNUSUAL FLORAL ORGANS* gene of *Arabidopsis* thaliana is an F-box protein required for normal patterning and growth in the floral meristem. Plant J. *20*, 433–445.

Schwarz, D.S., Hutvagner, G., Du, T., Xu, Z., Aronin, N., and Zamore, P.D. (2003). Asymmetry in the assembly of the RNAi enzyme complex. Cell *115*, 199–208.

Stark, A., Brennecke, J., Russell, R.B., and Cohen, S.M. (2003). Identification of *Drosophila* MicroRNA Targets. PLoS Biol *1*, E60.

Takahashi, H., Yamazaki, M., Sasakura, N., Watanabe, A., Leustek, T., Engler, J.A., Engler, G., Van Montagu, M., and Saito, K. (1997). Regulation of sulfur assimilation in higher plants: a sulfate transporter induced in sulfate-starved roots plays a central role in *Arabidopsis thaliana*. Proc. Natl. Acad. Sci. USA *94*, 11102–11107.

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.

Ulmasov, T., Hagen, G., and Guilfoyle, T.J. (1999). Dimerization and DNA binding of auxin response factors. Plant J. *19*, 309–319.

Vance, V.B. (1991). Replication of potato virus X RNA is altered in coinfections with potato virus Y. Virology *182*, 486–494.

Vaucheret, H., Vazquez, F., Crete, P., and Bartel, D.P. (2004). The action of *ARGONAUTE1* in the miRNA pathway and its regulation by the miRNA pathway are crucial for plant development. Genes Dev. *18*, 1187–1197.

Vazquez, F., Gasciolli, V., Crete, P., and Vaucheret, H. (2004). The nuclear dsRNA binding protein HYL1 is required for microRNA accumulation and plant development, but not posttranscriptional transgene silencing. Curr. Biol. *14*, 346–351.

Vierstra, R.D. (2003). The ubiquitin/26S proteasome pathway, the complex last chapter in the life of many plant proteins. Trends Plant Sci. *8*, 135–142.

Wilkinson, M.D., and Haughn, G.W. (1995). *UNUSUAL FLORAL ORGANS* controls meristem identity and organ primordia fate in *Arabidopsis*. Plant Cell *7*, 1485–1499.

Xie, Q., Frugis, G., Colgan, D., and Chua, N.H. (2000). Arabidopsis *NAC1* transduces auxin signal downstream of *TIR1* to promote lateral root development. Genes Dev. *14*, 3024–3036.

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.

Xie, Z., Johansen, L.K., Gustafson, A.M., Kasschau, K.D., Lellis, A.D., Zilberman, D., Jacobsen, S.E., and Carrington, J.C. (2004). Genetic and functional diversification of small RNA pathways in plants. PLoS Biol *2*, E104.

Yekta, S., Shih, I.H., and Bartel, D.P. (2004). MicroRNA-directed cleavage of *HOXB8* mRNA. Science *304*, 594–596.