Expression of Arabidopsis *MIRNA* Genes[1][w]

Zhixin Xie, Edwards Allen, Noah Fahlgren, Adam Calamar, Scott A. Givan, and James C. Carrington*

Center for Gene Research and Biotechnology, and Department of Botany and Plant Pathology, Oregon State University, Corvallis, Oregon 97331

MicroRNAs (miRNAs) are approximately 21-nucleotide noncoding RNAs that regulate target transcripts in plants and animals. In addition to miRNAs, plants contain several classes of endogenous small interfering RNAs (siRNAs) involved in target gene regulation and epigenetic silencing. Small RNA libraries were constructed from wild-type Arabidopsis (*Arabidopsis thaliana*) and mutant plants (*rdr2* and *dcl3*) that were genetically enriched for miRNAs, and a computational procedure was developed to identify candidate miRNAs. Thirty-eight distinct miRNAs corresponding to 22 families were represented in the libraries. Using a 5′ rapid amplification of cDNA ends procedure, the transcription start sites for 63 miRNA primary transcripts from 52 *MIRNA* loci (99 loci tested) were mapped, revealing features consistent with an RNA polymerase II mechanism of transcription. Ten loci (19%) yielded transcripts from multiple start sites. A canonical TATA box motif was identified upstream of the major start site at 45 (86%) of the mapped *MIRNA* loci. The 5′-mapping data were combined with miRNA cloning and 3′-PCR data to definitively validate expression of at least 73 *MIRNA* genes. These data provide a molecular basis to explore regulatory mechanisms of miRNA expression in plants.

MicroRNAs (miRNAs) are approximately 21-nt noncoding RNAs that posttranscriptionally regulate expression of target genes in plants and animals (Bartel, 2004). Mature miRNAs are generated through multiple processing steps from primary transcripts (pri-miRNA) that contain imperfect foldback structures. In animals, *MIRNA* genes are transcribed by RNA polymerase II (pol II; Bracht et al., 2004; Cai et al., 2004; Lee et al., 2004), yielding a pri-miRNA that is processed initially by the nuclear RNaseIII-like enzyme Drosha (Lee et al., 2003). The resulting pre-miRNA transcripts are transported to the cytoplasm and processed by Dicer to yield mature miRNAs (Lee et al., 2002). Less is known about the miRNA biogenesis pathway in plants, although most or all miRNAs require Dicer-like1 (DCL1; Park et al., 2002; Reinhart et al., 2002). The lack of a Drosha ortholog in plants and the finding that DCL1 functions at multiple steps during biogenesis of miR163 suggest that plant miRNA biogenesis may differ somewhat from animals (Kurihara and Watanabe, 2004). miRNAs in both animals and plants incorporate into an effector complex known as the RNA-induced silencing complex and guide either translation-associated repression or cleavage of target mRNAs (Bartel, 2004).

Computational and molecular cloning strategies revealed nearly 100 potential *MIRNA* genes in the Arabidopsis (*Arabidopsis thaliana*) genome (Llave et al., 2002a; Mette et al., 2002; Park et al., 2002; Reinhart et al., 2002; Palatnik et al., 2003; Jones-Rhoades and Bartel, 2004; Sunkar and Zhu, 2004; Wang et al., 2004). These miRNAs target mRNAs encoding proteins that include a variety of transcription factors involved in development, miRNA/small interfering RNA (siRNA) metabolic or effector components (DCL1, Argonaute1 [AGO1], and AGO2), components of the SCF complex involved in ubiquitin-mediated protein degradation, several other classes of metabolic and stress-related factors, as well as trans-acting siRNA (ta-siRNA) primary transcripts (Llave et al., 2002b; Park et al., 2002; Rhoades et al., 2002; Aukerman and Sakai, 2003; Emery et al., 2003; Kasschau et al., 2003; Palatnik et al., 2003; Tang et al., 2003; Xie et al., 2003; Achard et al., 2004; Allen et al., 2004, 2005; Chen, 2004; Jones-Rhoades and Bartel, 2004; Laufs et al., 2004; Mallory et al., 2004a; Sunkar and Zhu, 2004; Vaucheret et al., 2004; Vazquez et al., 2004). Based on tissue distribution and limited in situ expression data, most plant miRNAs are likely regulated during development (Chen, 2004; Juarez et al., 2004; Kidner and Martienssen, 2004; Parizotto et al., 2004). Overexpression or knockout of *MIRNA* genes, or expression of *MIRNA* genes outside of their normal domains, can lead to severe developmental defects (Aukerman and Sakai, 2003; Emery et al., 2003; Palatnik et al., 2003; Achard et al., 2004; Chen, 2004; Juarez et al., 2004; Kidner and Martienssen, 2004; Laufs et al., 2004; Mallory et al., 2004a, 2004b; McHale and Koning, 2004; Zhong and Ye, 2004). Understanding the mechanisms governing *MIRNA* gene expression patterns is necessary, therefore, to understand miRNA-mediated regulatory pathways and networks.

In this study, new Arabidopsis miRNAs were identified or validated by a computationally assisted

[1] This work was supported by the National Science Foundation (grant no. MCB–0209836), the National Institutes of Health (grant no. AI43288), and the U.S. Department of Agriculture (grant no. 2005–35319–15280).

* Corresponding author; e-mail carrington@cgrb.oregonstate.edu; fax 541–737–3045.

[w] The online version of this article contains Web-only data.

Article, publication date, and citation information can be found at www.plantphysiol.org/cgi/doi/10.1104/pp.105.062943.

cloning approach and the use of miRNA-enriched mutants. Features associated with transcription initiation of over one-half of all known Arabidopsis *MIRNA* genes were analyzed, revealing start sites, core promoter, and other properties that were consistent with a pol II mechanism of gene expression.

RESULTS AND DISCUSSION

Identification and Validation of Arabidopsis miRNAs

Several small RNA libraries were constructed from wild-type (Columbia [Col]-0) Arabidopsis seedling and inflorescence tissues, and from aerial tissues of *jaw-D* plants that overexpress miR-JAW (miR319; Palatnik et al., 2003). Among the 2,357 sequences analyzed collectively from these libraries, only 32.7% corresponded to known or subsequently validated miRNA families. Most of the remaining small RNAs corresponded to diverse sets of endogenous small RNAs arising from sequences such as transposons, retroelements, simple sequence repeats, inverted duplications, rDNA genes, and other genic and intergenic sequences (Llave et al., 2002a; Xie et al., 2004). To genetically enrich for miRNAs, small RNA libraries were constructed from embryo, seedling, and inflorescence tissues of *rdr2-1* mutant plants and from seedlings of *dcl3-1* mutant plants. These plants contain relatively low levels of approximately 24-nt siRNAs from repeated sequences but maintain normal levels of miRNAs (Xie et al., 2004). Among 3,164 sequences analyzed from the *rdr2-1* and *dcl3-1* libraries, 70.5% corresponded to previously characterized miRNAs, representing a 2.2-fold overall enrichment relative to the wild-type libraries. Endogenous siRNAs from known repeat families (identified from RepBase) were reduced 43.9-fold in the mutant libraries. The majority of the remaining small RNAs corresponded to sequences from two RDR2-independent small RNA-generating loci, or from rDNA genes. Unique miRNA and endogenous siRNA sequences from all libraries are available in the Arabidopsis Small RNA Project (ASRP) database (http://asrp.cgrb.oregonstate.edu/; Gustafson et al., 2005).

To identify new miRNAs in the cloned libraries, small RNA sequences were subjected to a series of six computational filters (Fig. 1A). The filters were designed using consensus properties of a founder set of published, validated Arabidopsis miRNAs with codes within the range of miR156 to miR399 (excluding miR390 and miR391; Llave et al., 2002a; Mette et al., 2002; Park et al., 2002; Reinhart et al., 2002; Palatnik et al., 2003; Jones-Rhoades and Bartel, 2004; Sunkar and Zhu, 2004), and using consensus properties of miRNAs from plants and animals (Ambros et al., 2003; Griffiths-Jones et al., 2003). Among the 48 unique miRNA sequences from 92 loci (22 validated miRNA families) in the founder set, 34 miRNA sequences from 71 loci (19 families) were in the cloned database. The initial filters eliminated small RNA sequences deriving from structural RNA genes, other annotated genes, and repetitive loci identified by RepeatMasker (Fig. 1A). Sequences originating from loci that yielded multidirectional clusters of small RNAs, which is a hallmark of many siRNA-generating loci, were eliminated. Small RNAs that were not 20 to 22 nt in length, based on the cloned sequence, were removed. Small RNAs originating from loci that lacked the potential to form a miRNA precursor-like foldback structure, consisting of a stem in which 16 or more positions within the putative miRNA:miRNA* duplex region were paired, were excluded. To test sensitivity, the complete founder set of miRNAs was processed through these filters. All but three passed, corresponding to a false negative rate of 0.032. miR163 failed due to length (24 nt), and two *MIR166* loci (*c* and *d*) failed because six or more mispaired bases were located within the miRNA/miRNA* region in the predicted foldback. From the cloned dataset, a total of 103 small RNAs passed each filter (Fig. 1A). These did not correspond to 103 unique loci, however, as many miRNA-generating loci yield multiple processed forms that were offset by one or a few nucleotides. The final filter eliminated all sequences corresponding to founder miRNAs, which yielded 18 small RNAs (13 loci) as candidate new miRNAs (Fig. 1A; Supplemental Table I). This set included miR390, miR391, miR403, and miR447 (Fig. 1B). miR403 and miR390 were also identified in an independent small RNA library (Sunkar and Zhu, 2004). Six of the 18 small RNAs corresponded to a cluster of processing variants from the two *MIR390* loci.

Given the sensitivity of the computational filters using the founder set, a second set of published Arabidopsis sequences with miRNA designations were analyzed. This set, which has not been subjected to extensive experimental validation, includes all sequences with codes between miR400 and miR420 (Sunkar and Zhu, 2004; Wang et al., 2004), except miR403. In contrast to the founder set, most of these small RNAs failed at one or more computational steps. Six small RNAs (miR401, 405a–d, 407, 416) were identified as transposon derived, and 10 (miR401, 404, 406, 408, 413, 414, 417–420) failed the foldback prediction criteria. Given the high computational failure rate (0.84), which was 26-fold higher than the false negative rate of the founder set, it is likely that many of these are endogenous siRNAs and not bona fide miRNAs.

Candidate miRNAs from each of the 13 loci identified in the computational analysis (Fig. 1A; Supplemental Table I) were subjected to validation-blot assays using a series of Arabidopsis miRNA-defective (*dcl1*, *hyl1*, *hen1*, and *hst*) and siRNA-defective (*dcl2*, *dcl3*, *rdr1*, *rdr2*, *rdr6*, and *sgs3*) mutants (Reinhart et al., 2002; Kasschau et al., 2003; Jones-Rhoades and Bartel, 2004; Peragine et al., 2004; Vazquez et al., 2004; Xie et al., 2004; Allen et al., 2005). The previously validated miR173, *AtSN1*-derived siRNAs and ta-siRNA1511

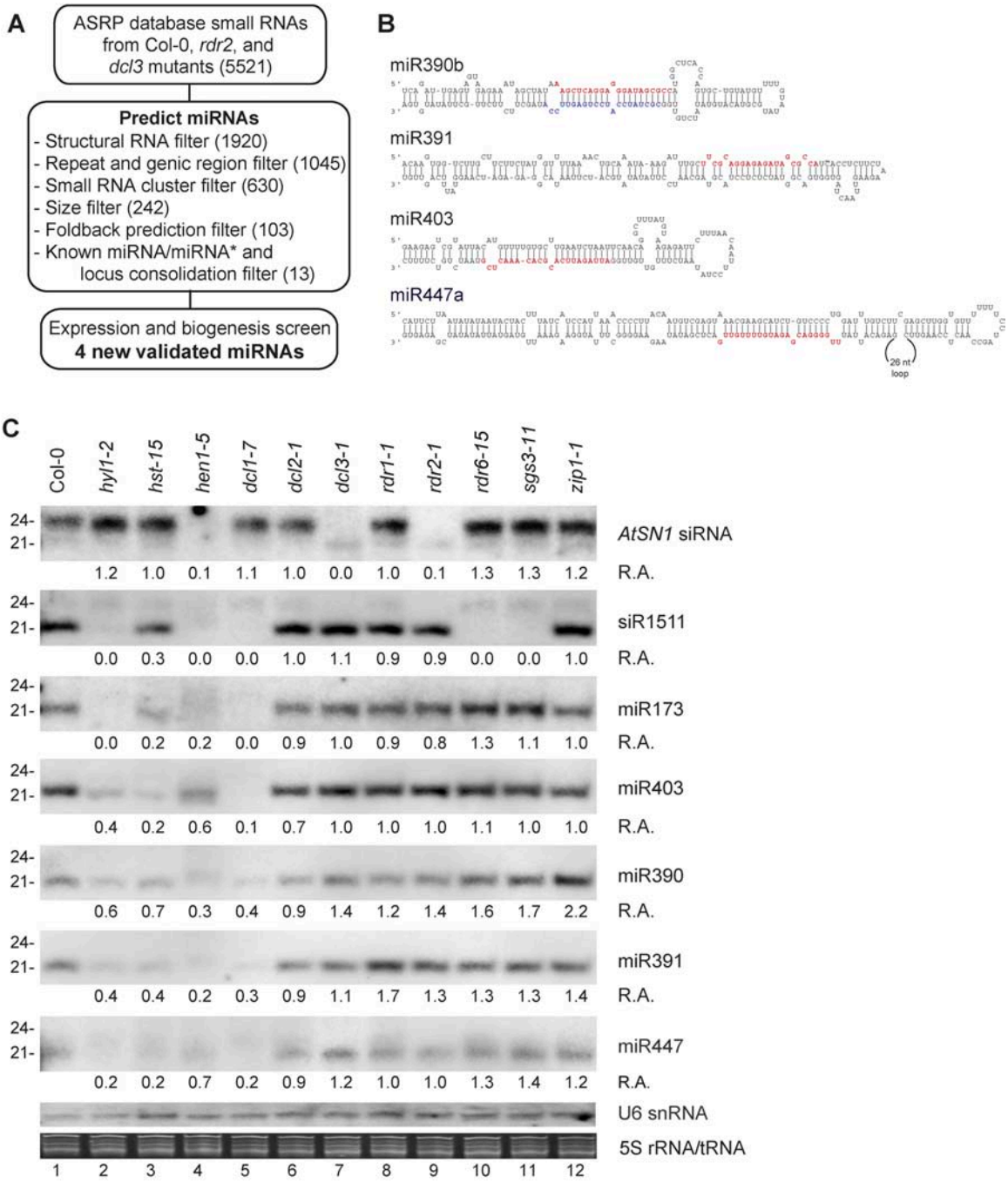

Figure 1. Cloning and identification of Arabidopsis miRNAs. A, Flowchart for identification miRNAs in cloned small RNA libraries. The number of small RNAs passing each filter is shown in parentheses. B, Predicted precursor foldback structure of miRNAs (red) validated in this study. C, Blot analysis of small RNAs in wild-type (Col-0) and mutant (*hyl1-2*, *hst-15*, *hen1-5*, *dcl1-7*, *dcl2-1*, *dcl3-1*, *rdr1-1*, *rdr2-1*, *rdr6-15*, *sgs3-11*, and *zip1-1*) plants. *AtSN1*-siRNAs and siR1511 represent two classes of endogenous siRNA. Relative accumulation (R.A.) of small RNAs in each mutant compared to wild-type Col-0 is shown.

were analyzed in parallel as controls. miR173, miR403, miR390, miR391, and miR447 each accumulated to relatively low levels in the *dcl1-7*, *hen1-5*, and *hyl1-2* mutants, but accumulated to normal or near-normal levels in the *dcl2-1*, *dcl3-1*, *rdr1-1*, *rdr2-1*, *rdr6-15*, *sgs3-11*, and *zip1-1* mutants (Fig. 1C). The *hst-15* mutant had a moderate effect on accumulation of all miRNAs (Fig. 1C), as shown previously (Park et al., 2005). Based on structural and biogenesis criteria, as well as target validation data (Allen et al., 2005), we conclude that miR390, miR391, miR403, and miR447 are bona fide miRNAs. Small RNAs from the remaining eight loci (Supplemental Table I) were not detected in blot assays and were not characterized further.

Table I. *Arabidopsis miRNA families*

	miRNA Family	Locus	Sequence[a]	ASRP Library[b]		Plant Species[c]	Target Family
				Col-0	*rdr2/dcl3*		
1	miR156	a-f	**U**GACAGAAGA**G**AG**U**GAGCAC	+	+	At, **Bn**, **Gm**, **Ha**, **Hv**, **Lj**, **Mt**, **Nt**, **Os**, **Pta**, Ptr, **Sb**, **Si**, **So**, **St**, **Vv**, **Zm**	SBP[d,e]
	miR156	g	**C**GACAGAAGA**G**AG**U**GAGCACA	−	−	At	
	miR156	h	**UU**GACAGAAGA**A**AG**A**GAGCAC	−	−	At	
	miR157	a-d	**UU**GACAGAAGA**U**AG**A**GAGCAC	−	+	At, Ptr	
2	miR158	a	**U**CCCAAAUGUAGACAAAGCA	+	−	At	
		b	**C**CCCAAAUGUAGACAAAGCA	−	−	At	
3	miR159	a	**U**UUGGA**U**UGAAGGGAGCUC**UA**	+	+	At, **Gm**, **Hv***, **Lj**, **Mt**, **Os**, **Pg***, **Ptr**, **So***, **Sb***, **Ta***, **Vv**, **Zm**	MYB[d,f,g]
	miR159	b	**U**UUGGA**U**UGAAGGGAGCUC**UU**	−	+	At	
	miR159	c	**U**UUGGA**U**UGAAGGGAGCUC**CU**	−	−	At	
	miR319	a-b	UUGGA**C**UGAAGGGAGCUC**CU**	+	+	At, **Bo**, **Gm**, **Lt**, **Os**, Ptr, **Ta**	TCP[g]
	miR319	c	UUGGA**C**UGAAGGGAGCUC**UU**	−	−	At, Os	
4	miR160	a-c	UGCCUGGCUCCCUGUAUGCCA	+	+	At, **Gm**, Os, Ptr, **Tt**, **Zm**	ARF[d,e]
5	miR161.1	a	UUGAAAGUGACUA**CAUCGGGG**	+	+	At	PPR[d,h]
	miR161.2	a	**UCAAUGCA**UUGAAAGUGACUA	+	+	At	
6	miR162	a-b	UCGAUAAACCUCUGCAUCCAG	+	+	At, **Gm**, **Ll**, **Mt**, **Os**, Ptr, **Vv**	DCL1[i]
7	miR163	a	UUGAAGAGGACUUGGAACUUCGAU	+	−	At	SAMT[h]
8	miR164	a-b	UGGAGAAGCAGGGCACGUGC**A**	−	+	At, **Pb**, **Ta**	NAC[dejk]
	miR164	c	UGGAGAAGCAGGGCACGUGC**G**	+	+	At	
9	miR165	a-b	UCGGACCAGGCUUCAU**C**CCCC	−	+	At, **Hc**, Ptr	HD-ZIPIII[lm]
	miR166	a-g	UCGGACCAGGCUUCAU**U**CCCC	+	+	At, **Gm**, **Hv**, **In***, **Mt**, Os, Ptr, **Sb**, **Zm**	
10	miR167	a-b	U**G**AAGCUGCCAGCAUGAUCU**A**	+	+	At, **Gm**, **Os**, **Pc***, Ptr, **Zm**	ARF[de]
	miR167	c	U**U**AAGCUGCCAGCAUGAUCU**U**	−	−	At	
	miR167	d	U**G**AAGCUGCCAGCAUGAUCU**GG**	+	+	**At**, **Gm**, **In**, **Ptr**, **So**	
11	miR168	a-b	UCGCUUGGUGCAGGUCGGGAA	+	+	**At**, **Bp**, **Gm**, **Ht**, **Hv**, **Le**, **Os**, **Ptr**, **Sb**, **So**, **St**, **Vv**, **Zm**	AGO1[dn]
12	miR169	a	**C**AGCCAAGGAUGACUUGCC**GA**	+	+	At, **Gm**, **Os**, **Ptr**	HAP2[o]
	miR169	b-c	**C**AGCCAAGGAUGACUUGCC**GG**	+	+	At, **Gm**, Os, Ptr, **Zm**	
	miR169	d-g	**UG**AGCCAAGGAUGACUUGCC**G**	+	+	At, Ptr	
	miR169	h-n	**U**AGCCAAGGAUGACUUGCC**UG**	+	+	At, **Ls**, **Os**, **Pb**, Ptr, **Sb**, **So**, **Ta**	
13	miR170	a	UGAUUGAGCCG**U**G**U**CAAUAUC	−	+	At	SCR[dp]
	miR171	a	UGAUUGAGCCG**C**G**C**CAAUAUC	+	+	At, Os, Ptr, Ta, Zm	
	miR171.2	b-c	UUGAGCCG**U**G**C**CAAUAUC**ACG**	+	−	At, Os, Ptr, **Ta**, **Zm**	
	miR171.1	c	UGAUUGAGCCG**U**G**C**CAAUAUC	−	+	At, **Gm**, **Hc**, **Hv**, Os, Ptr, **Ta**, **Zm**	
14	miR172	a-b	**A**GAAUCUUGAUGAUGCUGCA**U**	−	+	At, **Gm**, **Le**, Os, Ptr, **St**	AP2[eqr]
	miR172	c-d	**A**GAAUCUUGAUGAUGCUGCA**G**	+	−	At, **Cs**	
	miR172	e	**G**GAAUCUUGAUGAUGCUGCA**U**	−	+	At, Os, Ptr	
15	miR173	a	UUCGCUUGCAGAGAGAAAUCAC	−	+	At	TAS1, TAS2[s]
16	miR390	a-b	**AAG**C**U**CAGGAG**G**GAUAGCGCC	+	+	At, **Os**, Ptr, **St**, **Zm**	TAS3[s]
	miR391	a	**UU**C**G**CAGGAG**A**GAUAGCGCC**A**	−	+	At	
17	miR393	a-b	UCCAAAGGGAUCGCAUUGAUC	−	−	At, **Os**, **Ptr**	TIR1/F-box[o] bHLH[o]
18	miR394	a-b	UUUGGCAUUCUGUCCACCUCC	−	−	At, **Gm**, Os, Ptr, **Rp**	F-box[o]
19	miR395	a, d-e	CUGAAGUGUUUGGGGG**AA**CUC	−	−	At, **Gm**, **Os**, Ptr, **Ta**	ATPS[o], AST[s]
	miR395	b-c, f	CUGAAGUGUUUGGGGG**G**ACUC	−	−	At	
20	miR396	a	UUCCACAGCUUUCUUGAACU**G**	−	+	At, **Bv**, **Gm**, **Mc**, **Os**, **Ptr**, **So**, **St**, **Zm**	GRF[o]
	miR396	b	UUCCACAGCUUUCUUGAACU**U**	−	−	At, **Bn**, **Gm**, **Mc**, Os, Ptr, **St**	
21	miR397	a	UCAUUGAGUGCA**G**CGUUGAUG	−	+	At, **Hv**, Os, Ptr	Laccase[o]
	miR397	b	UCAUUGAGUGCA**U**CGUUGAUG	−	−	At	
22	miR398	a	UGUGUUCUCAGGUCACCCCU**U**	−	−	At, **Cs**, **Gm**, **Lj**, **Mt**, Os, Ptr	CSD[o]
	miR398	b-c	UGUGUUCUCAGGUCACCCCU**G**	−	+	At, **Gm**, **Ha**, **Ls**, **Mt**, **Nb**, **Os**, **Zm***	CytC oxidase[o]
23	miR399	a	UGCCAAAGGAGA**U**UUGCC**CU**G	−	−	At	E2-UBC[s]
	miR399	b, c	UGCCAAAGGAGA**G**UUGCC**CU**G	−	+	At, **Mt**, Os, **Ptr**, **Sb**	

(*Table continues on following page.*)

Table I. (*Continued from previous page.*)

	miRNA Family	Locus	Sequence[a]	ASRP Library[b] Col-0	ASRP Library[b] *rdr2/dcl3*	Plant Species[c]	Target Family
	miR399	d	UGCCAAAGGAGA**U**UUGCC**CC**G	−	−	At, Os	
	miR399	e	UGCCAAAGGAGA**U**UUGCC**UC**G	−	−	At	
	miR399	f	UGCCAAAGGAGA**U**UUGCC**CG**G	−	−	At, Os	
24	miR403	a	a**U**UAGAUUCACGCACAAACUCG	+	−	At, **Ptr**	AGO2[s]
25	miR447	a-b	UUGGGGACGA**G**AUGUUUUGUUG	−	+	At	2PGK[s]
	miR447	c	UUGGGGACGA**C**AUCUUUUGUUG	−	−		

[a]miRNAs are grouped by related families, with differences among families marked in bold. [b]Col-0 libraries included Col-0 and *jaw-d* sequences. [c]Presence of miRNA in genomic sequence is indicated in regular text, EST sequences are in bold, and sequences with 1 to 2 base changes from the Arabidopsis sequence are indicated by an asterisk. See Supplemental Table IV for plant species abbreviations. [d]Vazquez et al. (2004). [e]Kasschau et al. (2003). [f]Achard et al. (2004). [g]Palatnik et al. (2003). [h]Allen et al. (2004). [i]Xie et al. (2003). [j]Mallory et al. (2004a). [k]Laufs et al. (2004). [l]Tang et al. (2003). [m]Emery et al. (2003). [n]Vaucheret et al. (2004). [o]Jones-Rhoades and Bartel (2004). [p]Llave et al. (2002b). [q]Aukerman and Sakai (2003). [r]Chen (2004). [s]Allen et al. (2005).

The targets for miR390, miR403, and miR447 were recently predicted and validated (Allen et al., 2005). Genes encoding AGO2 (At1g31280) and a 2-phosphoglycerate kinase (2PGK, At5g60760) were validated as targets of miR403 and miR447, respectively (Table I; Allen et al., 2005). Interestingly, miR390 was shown to target primary transcripts from a ta-siRNA-generating locus (*TAS3*; Table I; Allen et al., 2005), setting the phase for subsequent processing of pre-ta-siRNAs through the RDR6/SGS3-dependent pathway (Allen et al., 2005).

miR390 and miR391 are related miRNAs that differ by 5 nt, whereas miR403 and miR447 are distinct from all other known miRNAs. If miR390 and miR391 are assigned to the same family, then Arabidopsis contains at least 25 experimentally validated families of miRNAs encoded by up to 99 genes (Table I). Among these families, 19 are conserved between dicots and monocots. One family (miR403) is conserved among families within dicots, and five families (miR158, miR161, miR163, miR173, and miR447) have been identified only in Arabidopsis.

Arabidopsis miRNA Primary Transcripts

To determine if a reference set of three Arabidopsis *MIRNA* gene transcripts contains 5′-cap structures typical of RNA pol II transcripts, a series of RNA ligase-mediated (RLM)-5′RACE reactions were done using poly(A^+)-selected RNA that was pretreated with either calf intestine phosphatase plus tobacco acid pyrophosphatase (CIP + TAP) or buffer alone. Only transcripts containing a 5′ cap should ligate to adapters, and subsequently amplify by PCR, following CIP + TAP treatment. Transcripts lacking a cap should ligate and amplify only from the sample treated with buffer alone. As controls, capped *Scarecrow-like6-IV* (*SCL6-IV*, At4g00150) mRNA and miR171-guided 3′-cleavage product from *SCL6-IV* (containing a 5′ monophosphate) were analyzed using gene-specific primer sets (Fig. 2A; Llave et al., 2002b). CIP + TAP-dependent 5′RACE products of the predicted size, approximately 400 and approximately 1,110 bp, were detected using 5′-proximal and cleavage site-proximal primer sets, respectively (Fig. 2B, lanes 2 and 4). CIP + TAP-independent 5′RACE product was detected only using the cleavage site-proximal primer set (Fig. 2B, lanes 1 and 3). Using locus-specific primer sets for

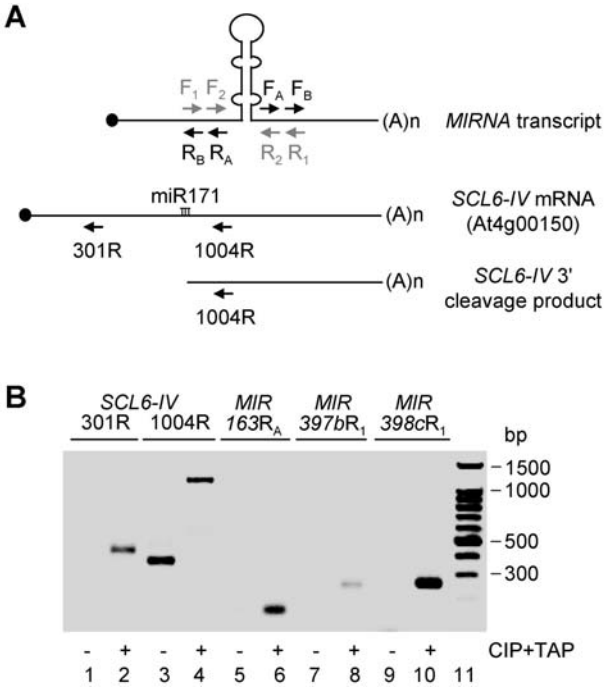

Figure 2. RLM-5′RACE on *MIRNA* transcripts. A, Schematic representation of a generic *MIRNA* transcript, and *SCL6-IV* mRNA (5′ cap-containing control) and miR171-guided cleavage product from *SCL6-IV* mRNA (noncapped control). The relative positions of oligonucleotide primers used in 5′RACE and 3′RACE reactions are shown, with the alternative primer sets shown in gray. B, RLM-5′RACE reactions using poly(A^+)-selected RNA. RNA was either pretreated with CIP + TAP (even-numbered lanes) or with buffer (odd-numbered lanes) prior to adaptor ligation. The 5′RACE products for *SCL6-IV* mRNA or internal cleavage product (lanes 1–4) and three *MIRNA* loci (lanes 5–10) were resolved on a 2% agarose gel. Gene-specific primers used in each reaction are indicated above each lane.

MIR163, *MIR397b*, and *MIR398c*, CIP + TAP-dependent products, but not CIP + TAP-independent products, were detected (Fig. 2B, lanes 5–10), indicating that the 5′ end of each miRNA transcript was capped. For 52 out of the 99 Arabidopsis *MIRNA* loci tested, 5′RACE products from poly(A^+)-selected and 5′-capped RNA were detected (see below and Supplemental Table II). Recently, a diverse set of putative miRNA transcripts from more than 30 plant species were identified from expressed sequence tag (EST) databases, suggesting that plant miRNA precursor transcripts are likely polyadenylated (Jones-Rhoades and Bartel, 2004). These data, combined with previous characterization of *MIR172b* and *MIR163* transcripts, indicate that plant *MIRNA* genes are transcribed by an RNA pol II mechanism. These data are also consistent with recent analyses of *MIRNA* gene transcripts from animal systems (Bracht et al., 2004; Cai et al., 2004; Lee et al., 2004).

Identification of Core Promoter Elements for Arabidopsis *MIRNA* Genes

For 52 of the 99 *MIRNA* genes tested, 5′RACE products were detected using locus-specific primers. In most cases (41 loci), positive 5′RACE products were detected as a uniform-sized fragment. At the remaining 10 loci, however, multiple 5′RACE products were detected. Nine of these (*MIR156a*, *MIR156c*, *MIR160a*, *MIR162a*, *MIR164a*, *MIR172a*, *MIR172c*, *MIR394a*, and *MIR447b*) gave rise to 5′RACE products of two distinct sizes. The other two, *MIR172b* and *MIR172e*, gave rise to three distinct 5′RACE products. Each PCR product from the 52 positive loci was cloned and sequenced, and transcription start sites were inferred based on the most abundant 5′ position represented among six or more clones randomly selected for sequencing. In cases where two 5′ positions were represented equally from one PCR size class, the extreme 5′ sequence was assigned as the start site. At 10 of the 11 loci (with the exception of *MIR172b*) for which multiple 5′RACE products were detected, alternative transcription start sites were identified. In the case of *MIR172b*, the three 5′RACE products corresponded to alternatively spliced transcripts that initiated at the same start site (Supplemental Fig. 1). Thus, 5′ ends representing 63 transcripts from 52 *MIRNA* loci were identified (Fig. 3). The vast majority (86%) of transcripts initiated with an adenosine, of which 93% were preceded by a pyrimidine (Fig. 3A). These characteristics are consistent with transcription by RNA pol II (Lorkovic et al., 2000; Shahmuradov et al., 2003).

Several characteristics of *MIR163* and *MIR172b* primary transcripts were reported in two previous studies (Aukerman and Sakai, 2003; Kurihara and Watanabe, 2004). The *MIR163* transcription start site mapped here is identical to the site identified previously (Kurihara and Watanabe, 2004). Splicing variants for both *MIR163* and *MIR172b* were reported (Aukerman and Sakai, 2003; Kurihara and Watanabe,

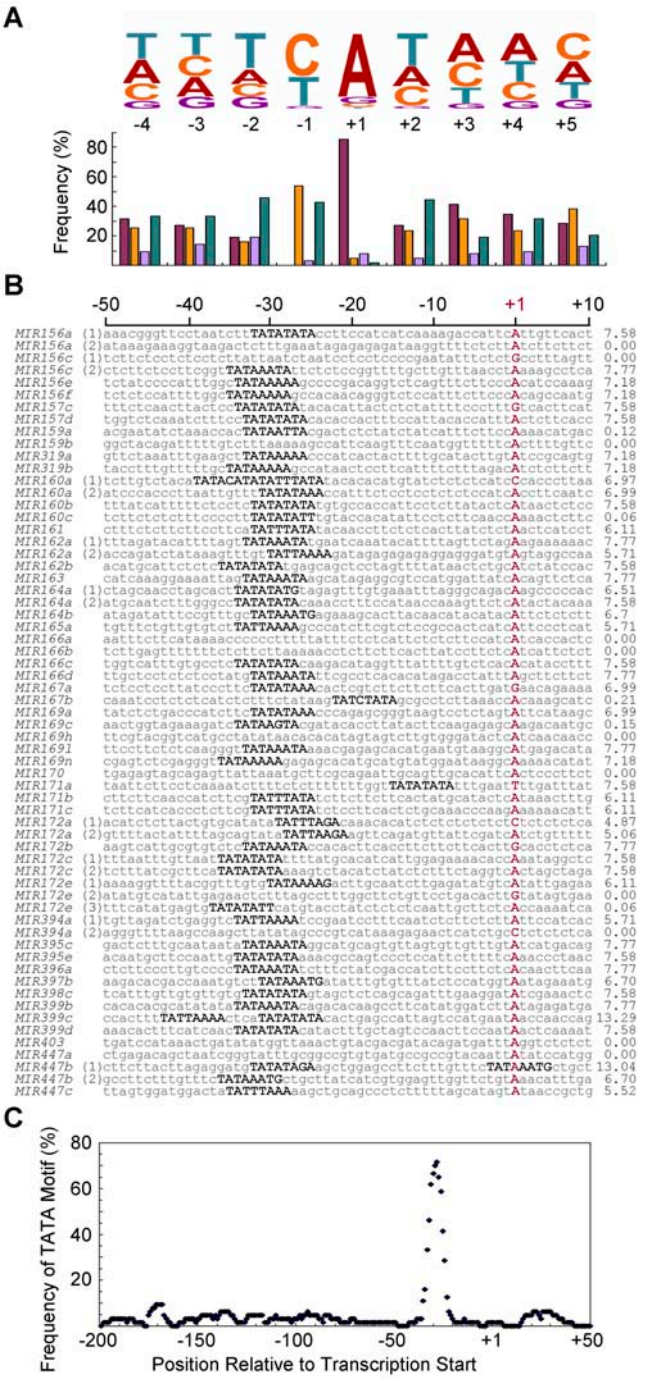

Figure 3. Transcription initiation sites of Arabidopsis *MIRNA* primary transcripts, and core promoter elements. A, Base composition at positions flanking *MIRNA* transcription initiation sites ($n = 63$). B, Genomic sequence of 60 nt flanking each of 63 *MIRNA* initiation sites (red letters) from 52 *MIRNA* loci. Putative TATA box-like motifs (bold) are indicated. MotifMatcher scores are given at the end of each sequence. C, Frequency of high-scoring TATA box-like motifs within a 250-nt (−200 to +50) context encompassing all mapped *MIRNA* transcript initiation sites. Frequency (%) was determined by count of sequences with TATA box-like motif within a single-nucleotide scrolling window.

Table II. *Position-weight matrix for conserved TATA box-like motif*

Nucleotide	Position (5′ to 3′)							
	1	2	3	4	5	6	7	8
A	0.0005	0.9988	0.0005	0.7747	0.4895	0.9988	0.2857	0.8766
C	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004
G	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0613	0.1224
T	0.9990	0.0006	0.9990	0.2247	0.5100	0.0006	0.6526	0.0006

2004). We identified two introns (117 nt and 199 nt) separated by 41 nt upstream of the predicted foldback structure in the *MIR172b* transcript (Supplemental Fig. 1). Only one intron was reported in this region by Aukerman and Sakai (2003). The *MIR163* intron was not detected in this study, as 5′RACE primers corresponded to a sequence upstream of the intron. An intron was also identified in one of the two *MIR156a* transcripts (Supplemental Fig. 1). In each case, the intron began with 5′-GU and ended with AG-3′, as is

Locus	5′ RACE	3′ RACE	Cloned in ASRP	Other Refs.
MIR156a	Yes	nt	228	[11]
MIR156b	No	No		
MIR156c	Yes	nt		
MIR156d	No	nt		
MIR156e	Yes	nt		
MIR156f	Yes	nt		
MIR156g	No	No		
MIR156h	No	No		
MIR157a	No	Yes	9(1)	
MIR157b	No	Yes		
MIR157c	Yes	nt		
MIR157d	Yes	nt		
MIR158a	No	nt	18(3)	[11]
MIR158b	No	No		
MIR159a	Yes	nt	225	[8-12]
MIR159b	Yes	nt	7	
MIR159c	No	No		
MIR319a	Yes	nt	5	
MIR319b	Yes	nt		
MIR319c	No	nt		
MIR160a	Yes	nt	103(1)	[11]
MIR160b	Yes	nt		
MIR160c	Yes	nt		
MIR161	Yes	nt	334	[1,6,11]
MIR162a	Yes	Yes	4	[11]
MIR162b	Yes	Yes		
MIR163	Yes	nt	1	[1,5,10,11]
MIR164a	Yes	nt	2	[11]
MIR164b	Yes	nt		
MIR164c	No	nt	0(2)	
MIR165a	Yes	nt	30	[11]
MIR165b	No	No		
MIR166a	Yes	nt	307(1)	
MIR166b	Yes	nt		
MIR166c	Yes	nt		
MIR166d	Yes	nt		
MIR166e	No	Yes		
MIR166f	No	Yes		
MIR166g	No	No		
MIR167a	Yes	nt	173	[6,10, 11,13]
MIR167b	Yes	nt		
MIR167c	No	No		
MIR167d	No	nt	7	
MIR168a	No	nt	22(7)	[11,13]
MIR168b	No	No		
MIR169a	Yes	nt	619(1)	[11]
MIR169b	No	nt	128(1)	
MIR169c	Yes	nt		
MIR169d	No	No	22	
MIR169e	No	No		
MIR169f	No	No	22	[11]
MIR169g	No	No		
MIR169h	Yes	nt	458	
MIR169i	No	nt		
MIR169j	No	No		
MIR169k	No	No		
MIR169l	Yes	nt		
MIR169m	No	nt		
MIR169n	Yes	nt		
MIR170	Yes	nt	1	[6,7, 11,12]
MIR171a	Yes	nt	35	
MIR171b	Yes	nt	2	
MIR171c	Yes	nt		
MIR172a	Yes	nt	0(2)	[2,3,10]
MIR172e	Yes	nt		
MIR172b	Yes	nt		
MIR172c	Yes	nt	1	
MIR172d	No	No		
MIR173	No	nt	1(1)	[10]
MIR390a	No	nt	122(2)	[12]
MIR390b	No	nt		
MIR391	No	nt	7(1)	
MIR393a	No	No		[4,12]
MIR393b	No	No		
MIR394a	Yes	nt		[4]
MIR394b	No	No		
MIR395a	No	No		[4]
MIR395b	No	No		
MIR395c	Yes	nt		
MIR395d	No	No		
MIR395e	Yes	nt		
MIR395f	No	No		
MIR396a	Yes	nt	0(1)	[4]
MIR396b	No	No		
MIR397a	No	nt	1	[4,12]
MIR397b	Yes	nt		
MIR398a	No	nt	4	[4,12]
MIR398b	No	Yes		
MIR398c	Yes	nt		
MIR399a	No	No	1	[4,12]
MIR399b	Yes	nt		
MIR399c	Yes	nt		
MIR399d	Yes	nt		
MIR399e	No	No		
MIR399f	No	nt		
MIR403	Yes	Yes	1	[12]
MIR447a	Yes	nt	1	
MIR447b	Yes	nt		
MIR447c	Yes	nt		

Figure 4. Locus-specific expression of 99 predicted *MIRNA* genes encoding validated miRNAs in Arabidopsis. Expression of a specific locus was considered definitive (dark green shading) if a primary transcript was detected by 5′RACE or 3′RACE, or a unique miRNA sequence was cloned or amplified from the ASRP library described here (gray shading with total clones sequenced) or from another published library (Other Refs.). The number of clones corresponding to a specific miRNA or miRNA* (in parentheses) sequence in the ASRP database is shown. Sequences that were detected only in other studies are indicated by orange in the 3′RACE and references columns. Loci for which data support expression from more than one possible gene are indicated by light green shading. nt, Not tested. References cited are as follows: 1, Allen et al. (2004); 2, Aukerman and Sakai (2003); 3, Chen (2004); 4, Jones-Rhoades and Bartel (2004); 5, Kurihara and Watanabe (2004); 6, Llave et al. (2002a); 7, Llave et al. (2002b); 8, Mette et al. (2002); 9, Palatnik et al. (2003); 10, Park et al. (2002); 11, Reinhart et al. (2002); 12, Sunkar and Zhu (2004); and 13, Arabidopsis EST clones were identified for *MIR167d* (GenBank accession no. AU239920) and *MIR168a* (H77158).

typical of group III introns that are commonly found in pre-mRNAs of higher plants (Lorkovic et al., 2000).

To identify conserved motifs flanking the initiation sites at each mapped locus, a 60-bp genomic segment (−50 to +10 relative to the start site) was computationally analyzed using BioProspector (Liu et al., 2004). An 8-nt TATA box-like sequence was identified as a conserved motif in this region (Table II). This motif was detected upstream from 52 (83%) of the mapped transcription start sites (Fig. 3B). To determine if the high frequency occurrence of the TATA box-like sequence was uniquely associated with this specific region, we examined the distribution of the TATA box-like sequence in an extended upstream region (−200 to +50) using MotifMatcher. The TATA box-like sequence centered at consensus position −29 from the start site (Fig. 3C), which is entirely consistent with TATA box motifs located in protein-coding genes (Patikoglou et al., 1999; Shahmuradov et al., 2003). We conclude, therefore, that most or all of these motifs correspond to authentic TATA box sequences within core promoters of *MIRNA* genes.

Expression of Arabidopsis *MIRNA* Genes

Despite repeated attempts with multiple primer sets, 5′RACE products were detected from only 53% of *MIRNA* genes tested (Fig. 4). This may have been due to low levels of expression of some *MIRNA* genes in tissues analyzed or lack of expression of some loci predicted to be *MIRNA* genes. It is also possible that some primer sets were inadvertently designed within intron sequences or that some miRNA sequences derive from non-5′ positions within polycistronic primary transcripts. To develop a more comprehensive account of expression of Arabidopsis *MIRNA* genes, informatic and experimental approaches were taken. In the informatic strategy, the ASRP database was scanned for miRNA or miRNA* sequences corresponding to loci with negative 5′RACE results (Gustafson et al., 2005). Unique miRNA or miRNA* sequences from *MIR156d*, *MIR158a*, *MIR164c*, *MIR167d*, *MIR168a*, *MIR169b*, *MIR169i*, *MIR169m*, *MIR173*, *MIR390b*, *MIR391*, *MIR397a*, and *MIR403* loci were each represented in the database (Fig. 4). For *MIR168a*, additional evidence confirming expression came from a locus-specific EST clone (GenBank accession no. H77158). In addition, unique miRNA sequences specific to *MIR319c*, *MIR390a*, *MIR398a*, and *MIR399f* were each represented in an independent Arabidopsis small RNA library (Fig. 4; Sunkar and Zhu, 2004).

For the remaining predicted *MIRNA* genes, locus-specific primers were designed to amplify sequences immediately downstream of the precursor foldback sequence using a 3′RACE procedure. Three 5′RACE-positive *MIRNA* loci (*MIR162a*, *MIR162b*, and *MIR403*) were also included in the 3′RACE analysis as controls. Positive results were obtained for *MIR157a*, *MIR157b*, *MIR166e*, *MIR166f*, and *MIR398b* as well as for the three control loci (Fig. 4). For the miR393 family and a subset of members in the miR169 family (*MIR169d–g*), neither 5′RACE nor 3′RACE yielded positive results (Fig. 4), although expression from at least one locus was inferred based on detection of a sequence in at least one small RNA library (Jones-Rhoades and Bartel, 2004; Sunkar and Zhu, 2004; Gustafson et al., 2005). Collectively, unambiguous data support the expression of at least 73 of the 99 Arabidopsis *MIRNA* loci (Fig. 4).

Each of the genes encoding miRNAs that are conserved between monocots and dicots are members of multigene families. For nearly all of these families, multiple genes are expressed and presumed functional. In some cases, the family variants encode miRNAs with diverged sequences. We propose two forces are driving the evolution of these multigene families. First, expansion of *MIRNA* gene families facilitates regulatory diversification through acquisition or derivation of distinct control elements (Hurles, 2004). Although the extent to which *MIRNA* gene duplication leads to novel spatial or temporal control of family members remains to be determined, genetic data from *MIR164* loci indicate divergence of regulatory specificity between closely related family members (Baker et al., 2005). And, second, family expansion provides the genes to generate novel miRNAs with unique target specificity, as proposed for the miR159/319 family and *MYB/TCP* gene targets (Palatnik et al., 2003). The finding that most members of conserved, multigene *MIRNA* families are expressed in Arabidopsis supports the idea that regulatory or functional diversification has occurred.

MATERIALS AND METHODS

Upon request, all novel materials described in this publication will be made available in a timely manner for noncommercial research purposes, subject to the requisite permission from any third-party owners of all or parts of the material. Obtaining any permission will be the responsibility of the requestor.

Cloning of Arabidopsis Small RNAs and miRNA Prediction

Extraction of low molecular weight RNA and library construction were done as described (Lau et al., 2001; Llave et al., 2002a). RNA was extracted from 3-d postgermination seedlings, embryos from developing siliques, aerial tissues including rosette leaves and apical meristems, or stage 1 to 12 enriched inflorescence from wild type Col-0, and *jaw-D*, *rdr2-1*, and *dcl3-1* mutants as described previously (Palatnik et al., 2003; Xie et al., 2004). Seedling libraries were constructed for Col-0, *rdr2-1*, and *dcl3-1*, embryo libraries for *rdr2-1*, aerial libraries for *jaw-D*, and inflorescence libraries for Col-0 and *rdr2-1*. The procedure for small RNA cloning and sequencing was described previously (Llave et al., 2002a). A total of 2,357 sequences were determined from wild-type or *jaw-D* libraries, and 3,164 sequences were determined from *rdr2-1* and *dcl3-1* libraries. miRNAs were predicted from the cloned database of sequences using a set of six computational filters. First, structural RNAs were identified by BLAST and eliminated. Second, small RNAs from repeated sequences identified using RepeatMasker (Jurka, 2000), or from predicted protein-coding genes and pseudogenes only, were removed. Third, a small RNA cluster filter was applied to remove small RNAs within 500 nt of another small RNA in the opposite orientation. The fourth filter removed any small RNAs outside the typical size range for miRNA (20–22 nt). Fifth, sequences from within a context that failed to conform to a set of consensus

characteristics of miRNA foldback structures were eliminated. The consensus criteria were (1) minimum of 16 paired bases within the miRNA:miRNA* duplex, (2) maximum predicted foldback size of 350 nt, (3) a requirement for the miRNA:miRNA* duplex to be predicted within a single foldback stem, and (4) three or fewer contiguous nonpaired bases. RNAFold in the Vienna RNA Package was used to predict potential duplexes containing the small RNA (Hofacker, 2003). The remaining small RNAs were compared to sequences of validated miRNAs by FASTA to identify previously characterized and unique candidate miRNAs.

Arabidopsis Mutants

Mutant lines for *dcl1-7*, *dcl2-1*, *dcl3-1*, *rdr1-1*, *rdr2-1*, *hen1-5*, *hyl1-2*, *rdr6-15*, *sgs3-11*, and *zip1-1* were described previously (Park et al., 2002; Allen et al., 2004; Peragine et al., 2004; Vazquez et al., 2004; Xie et al., 2004). The *hst-15* mutant was derived from a SALK_079290 T-DNA insertion line, which contains an insertion at position 1,584 from the start codon (Alonso et al., 2003).

Small RNA-Blot Analysis

Low molecular weight RNA (5 μg) from Arabidopsis (*Arabidopsis thaliana*) inflorescence tissue was used for miRNA- and endogenous siRNA-blot analysis as previously described (Allen et al., 2004). Probe for *AtSN1*-siRNA blot was described previously (Zilberman et al., 2003). DNA oligonucleotide probes specific for miR390 (5′-GGCGCTATCCCTCCTGAGCTT-3′) and miR391 (5′-TGGCGCTATCTCTCCTGCGAA-3′) were end-labeled with γ^{32}P-ATP using Optikinase (New England Biolabs, Beverly, MA) according to the manufacturer's directions. Locked nucleic acid-modified oligonucleotides (Exiqon, Vedbaek, Denmark) specific for siR1511 (5′-AAGTATCATCAT-TCGCTTGGA-3′), miR447 (5′-CAACAAAACATCTCGTCCCCAA-3′), and miR403 (5′-CGAGTTTGTGCGTGAATCTAAT-3′) were used to improve sensitivity of detection (Valoczi et al., 2004). The blots were also analyzed with an oligonucleotide probe specific to snRNA U6 (5′-TCATCCTTGCGCAGGGG-CCA-3′) as a loading control. Relative accumulation of small RNAs was determined using an InstantImager (Packard Bioscience, Boston, MA).

5′RACE Mapping of *MIRNA* Transcripts

Two Arabidopsis (Col-0) tissue preparations were used for RNA isolation: inflorescence tissues from 4-week-old plants grown under greenhouse conditions and 4-d-old seedlings grown on Murashige and Skoog media in a growth chamber. Total RNA was extracted using TRIzol reagent (Invitrogen, Carlsbad, CA), followed by column purification using an RNA/DNA midi kit (Qiagen, Valencia, CA). The extracts were subjected to two rounds of purification using oligo(dT) resin (Qiagen) for the enrichment of poly(A^+) RNA. Poly(A^+)-enriched RNA (125 ng/reaction) was first treated with CIP + TAP. The 5′ ends of *MIRNA* transcripts were then mapped by an RLM-5′RACE assay (Invitrogen). Complementary DNA (cDNA) was synthesized using random oligonucleotide hexamers as primers. A cDNA pool containing equal amounts of reaction product from each tissue was used as template in 5′RACE PCR with a primer specific to the RNA adaptor sequence and a locus-specific reverse primer. In general, a set of two gene-specific primers (R_A and R_B) were designed for each *MIRNA* locus based on sequences immediately upstream to the predicted foldback structure (Fig. 2A). In cases where the sequence context in this region did not allow designing primers with high specificity, or the size of resulting 5′RACE products was too small, an alternative set of primers (R_1 and R_2) were used (Fig. 2A; Supplemental Table III). The default annealing temperature in the touchdown PCR reaction was 65°C. For *MIRNA* loci yielding negative 5′RACE results after the second-round PCR, two additional PCR reactions with the nested primers were done with altered annealing temperatures. The PCR products from a positive 5′RACE reaction were gel purified and cloned. A minimum of six clones were sequenced for each PCR product. Sequences corresponding to transcript 5′ ends were deposited at GenBank with accession numbers listed in the supplemental materials (Supplemental Table II).

The RLM-5′RACE procedure was used to analyze the presence or absence of a cap structure on several miRNA primary transcripts. A capped mRNA [*SCL6-IV*] and a noncapped RNA (miR171-guided cleavage product of *SCL6-IV* mRNA) were used as control RNAs. Parallel RLM-5′RACE reactions were done using poly(A^+)-enriched RNA that was CIP + TAP treated (selective for 5′ ends that contain a 5′ cap) or buffer treated (selective for noncapped 5′ ends).

For some *MIRNA* primary transcripts, 3′RACE was done using poly(A^+)-enriched RNA. cDNA was synthesized using an adaptor-tagged oligo(dT) primer (Invitrogen). Gene-specific forward primers were designed for each locus tested, following the same procedure used for 5′RACE primer design (Fig. 2A). The identities of 3′RACE products were confirmed by sequencing. The sequences of all the locus-specific primers are listed in Supplemental Table III.

Computational Identification of Conserved Upstream Sequence Motifs

A 60-bp (−50 to +10) genomic sequence flanking the start site for 63 transcripts from 52 *MIRNA* loci was analyzed using BioProspector, a Gibbs sampling-based motif-finding program (Liu et al., 2004). Searches with a motif width of 6 to 8 nt were done. In all cases, TATA box-like sequences were identified as the only conserved motif. The presence of the conserved TATA box-like motif matrix (8-nt width) in each 60-bp genomic segment was checked using MotifMatcher, with up to three matches per segment allowed (Ao et al., 2004). The algorithm gives a score for placement of each TATA box-like sequence detected (Fig. 2B). These are log-odds-based scores calculated as $\ln[P(\text{observed}|\text{PWM})/P(\text{observed}|\text{background model})]$, where the numerator is the probability of the observed sequence according to the position weight matrix (PWM) representing the motif and the denominator is the probability of the sequence according to a simple Markov chain constructed by examining frequencies of nucleotide occurrences throughout a background sequence set (Ao et al., 2004). A second search by MotifMatcher was done using an extended upstream region (−200 to +50) to analyze the distribution of the putative TATA motif, with the 8-nt motif matrix generated by BioProspector as a sample motif. Up to three matches to the TATA box-like motif were allowed.

Sequence data from this article have been deposited with the EMBL/GenBank data libraries under accession numbers DQ063602 to DQ063665.

ACKNOWLEDGMENTS

We thank Scott Poethig for the *dcl1-7* allele in Col-0 background. We also thank Adam Gustafson, Daniel Smith, and Christopher Sullivan for invaluable assistance and advice with computational resources; April Wilken for assistance in PCR and cloning; and Mark Dasenko for sequencing.

Received March 15, 2005; revised May 22, 2005; accepted May 23, 2005; published July 22, 2005.

LITERATURE CITED

Achard P, Herr A, Baulcombe DC, Harberd NP (2004) Modulation of floral development by a gibberellin-regulated microRNA. Development **131:** 3357–3365

Allen E, Xie Z, Gustafson AM, Carrington JC (2005) MicroRNA-directed phasing during trans-acting siRNA biogenesis in plants. Cell **121:** 207–221

Allen E, Xie Z, Gustafson AM, Sung GH, Spatafora JW, Carrington JC (2004) Evolution of microRNA genes by inverted duplication of target gene sequences in *Arabidopsis thaliana*. Nat Genet **36:** 1282–1290

Alonso JM, Stepanova AN, Leisse TJ, Kim CJ, Chen H, Shinn P, Stevenson DK, Zimmerman J, Barajas P, Cheuk R, et al (2003) Genome-wide insertional mutagenesis of *Arabidopsis thaliana*. Science **301:** 653–657

Ambros V, Bartel B, Bartel DP, Burge CB, Carrington JC, Chen X, Dreyfuss G, Eddy SR, Griffiths-Jones S, Marshall M, et al (2003) A uniform system for microRNA annotation. RNA **9:** 277–279

Ao W, Gaudet J, Kent WJ, Muttumu S, Mango SE (2004) Environmentally induced foregut remodeling by PHA-4/FoxA and DAF-12/NHR. Science **305:** 1743–1746

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

Baker CC, Sieber P, Wellmer F, Meyerowitz EM (2005) The *early extra*

petals1 mutant uncovers a role for microRNA miR164c in regulating petal number in *Arabidopsis*. Curr Biol **15:** 303–315

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

Bracht J, Hunter S, Eachus R, Weeks P, Pasquinelli AE (2004) Trans-splicing and polyadenylation of let-7 microRNA primary transcripts. RNA **10:** 1586–1594

Cai X, Hagedorn CH, Cullen BR (2004) Human microRNAs are processed from capped, polyadenylated transcripts that can also function as mRNAs. RNA **10:** 1957–1966

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

Emery JF, Floyd SK, Alvarez J, Eshed Y, Hawker NP, Izhaki A, Baum SF, Bowman JL (2003) Radial patterning of *Arabidopsis* shoots by class III *HD-ZIP* and *KANADI* genes. Curr Biol **13:** 1768–1774

Griffiths-Jones S, Bateman A, Marshall M, Khanna A, Eddy SR (2003) Rfam: an RNA family database. Nucleic Acids Res **31:** 439–441

Gustafson AM, Allen E, Givan S, Smith D, Carrington JC, Kasschau KD (2005) ASRP: the *Arabidopsis* Small RNA Project Database. Nucleic Acids Res **33:** D637–D640

Hofacker IL (2003) Vienna RNA secondary structure server. Nucleic Acids Res **31:** 3429–3431

Hurles M (2004) Gene duplication: the genomic trade in spare parts. PLoS Biol **2:** E206

Jones-Rhoades MW, Bartel DP (2004) Computational identification of plant microRNAs and their targets, including a stress-induced miRNA. Mol Cell **14:** 787–799

Juarez MT, Kui JS, Thomas J, Heller BA, Timmermans MC (2004) MicroRNA-mediated repression of *rolled leaf1* specifies maize leaf polarity. Nature **428:** 84–88

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

Kasschau KD, Xie Z, Allen E, Llave C, Chapman EJ, Krizan KA, Carrington JC (2003) P1/HC-Pro, a viral suppressor of RNA silencing, interferes with *Arabidopsis* development and miRNA function. Dev Cell **4:** 205–217

Kidner CA, Martienssen RA (2004) Spatially restricted microRNA directs leaf polarity through ARGONAUTE1. Nature **428:** 81–84

Kurihara Y, Watanabe Y (2004) *Arabidopsis* micro-RNA biogenesis through Dicer-like 1 protein functions. Proc Natl Acad Sci USA **101:** 12753–12758

Lau NC, Lim EP, Weinstein EG, Bartel DP (2001) An abundant class of tiny RNAs with probable regulatory roles in *Caenorhabditis elegans*. Science **294:** 858–862

Laufs P, Peaucelle A, Morin H, Traas J (2004) MicroRNA regulation of the *CUC* genes is required for boundary size control in *Arabidopsis* meristems. Development **131:** 4311–4322

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

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

Lee Y, Kim M, Han J, Yeom KH, Lee S, Baek SH, Kim VN (2004) MicroRNA genes are transcribed by RNA polymerase II. EMBO J **23:** 4051–4060

Liu Y, Wei L, Batzoglou S, Brutlag DL, Liu JS, Liu XS (2004) A suite of web-based programs to search for transcriptional regulatory motifs. Nucleic Acids Res **32:** W204–W207

Llave C, Kasschau KD, Rector MA, Carrington JC (2002a) Endogenous and silencing-associated small RNAs in plants. Plant Cell **14:** 1605–1619

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

Lorkovic ZJ, Wieczorek Kirk DA, Lambermon MH, Filipowicz W (2000) Pre-mRNA splicing in higher plants. Trends Plant Sci **5:** 160–167

Mallory AC, Dugas DV, Bartel DP, Bartel B (2004a) MicroRNA regulation of NAC-domain targets is required for proper formation and separation of adjacent embryonic, vegetative, and floral organs. Curr Biol **14:** 1035–1046

Mallory AC, Reinhart BJ, Jones-Rhoades MW, Tang G, Zamore PD, Barton MK, Bartel DP (2004b) MicroRNA control of *PHABULOSA* in leaf development: importance of pairing to the microRNA 5′ region. EMBO J **23:** 3356–3364

McHale NA, Koning RE (2004) MicroRNA-directed cleavage of *Nicotiana sylvestris PHAVOLUTA* mRNA regulates the vascular cambium and structure of apical meristems. Plant Cell **16:** 1730–1740

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

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

Parizotto EA, Dunoyer P, Rahm N, Himber C, Voinnet O (2004) *In vivo* investigation of the transcription, processing, endonucleolytic activity, and functional relevance of the spatial distribution of a plant miRNA. Genes Dev **18:** 2237–2242

Park MY, Wu G, Gonzalez-Sulser A, Vaucheret H, Poethig RS (2005) Nuclear processing and export of microRNAs in *Arabidopsis*. Proc Natl Acad Sci USA **102:** 3691–3696

Park W, Li J, Song R, Messing J, 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

Patikoglou GA, Kim JL, Sun L, Yang SH, Kodadek T, Burley SK (1999) TATA element recognition by the TATA box-binding protein has been conserved throughout evolution. Genes Dev **13:** 3217–3230

Peragine A, Yoshikawa M, Wu G, Albrecht HL, Poethig RS (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

Reinhart BJ, Weinstein EG, Rhoades MW, Bartel B, Bartel DP (2002) MicroRNAs in plants. Genes Dev **16:** 1616–1626

Rhoades MW, Reinhart BJ, Lim LP, Burge CB, Bartel B, Bartel DP (2002) Prediction of plant microRNA targets. Cell **110:** 513–520

Shahmuradov IA, Gammerman AJ, Hancock JM, Bramley PM, Solovyev VV (2003) PlantProm: a database of plant promoter sequences. Nucleic Acids Res **31:** 114–117

Sunkar R, Zhu JK (2004) Novel and stress-regulated microRNAs and other small RNAs from Arabidopsis. Plant Cell **16:** 2001–2019

Tang G, Reinhart BJ, Bartel DP, Zamore PD (2003) A biochemical framework for RNA silencing in plants. Genes Dev **17:** 49–63

Valoczi A, Hornyik C, Varga N, Burgyan J, Kauppinen S, Havelda Z (2004) Sensitive and specific detection of microRNAs by northern blot analysis using LNA-modified oligonucleotide probes. Nucleic Acids Res **32:** e175

Vaucheret H, Vazquez F, Crete P, Bartel DP (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, 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

Wang XJ, Reyes JL, Chua NH, Gaasterland T (2004) Prediction and identification of *Arabidopsis thaliana* microRNAs and their mRNA targets. Genome Biol **5:** R65

Xie Z, Johansen LK, Gustafson AM, Kasschau KD, Lellis AD, Zilberman D, Jacobsen SE, Carrington JC (2004) Genetic and functional diversification of small RNA pathways in plants. PLoS Biol **2:** 642–652

Xie Z, Kasschau KD, Carrington JC (2003) Negative feedback regulation of *Dicer-Like1* in *Arabidopsis* by microRNA-guided mRNA degradation. Curr Biol **13:** 784–789

Zhong R, Ye ZH (2004) *Amphivasal vascular bundle 1*, a gain-of-function mutation of the *IFL1/REV* gene, is associated with alterations in the polarity of leaves, stems and carpels. Plant Cell Physiol **45:** 369–385

Zilberman D, Cao X, Jacobsen SE (2003) *ARGONAUTE4* control of locus-specific siRNA accumulation and DNA and histone methylation. Science **299:** 716–719