



MicroRNAs and other small RNAs enriched in the *Arabidopsis* RNA-dependent RNA polymerase-2 mutant

Cheng Lu, Karthik Kulkarni, Frédéric F. Souret, et al.

Genome Res. 2006 16: 1276-1288

Access the most recent version at doi:[10.1101/gr.5530106](https://doi.org/10.1101/gr.5530106)

References This article cites 53 articles, 24 of which can be accessed free at:
<http://genome.cshlp.org/content/16/10/1276.full.html#ref-list-1>

Open Access Freely available online through the *Genome Research* Open Access option.

License Freely available online through the Genome Research Open Access option.

Email Alerting Service Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).



To subscribe to *Genome Research* go to:
<https://genome.cshlp.org/subscriptions>

Copyright © 2006, Cold Spring Harbor Laboratory Press

MicroRNAs and other small RNAs enriched in the *Arabidopsis* RNA-dependent RNA polymerase-2 mutant

Cheng Lu,¹ Karthik Kulkarni,¹ Frédéric F. Souret,¹ Ramesh MuthuValliappan,¹ Shivakundan Singh Tej,¹ R. Scott Poethig,² Ian R. Henderson,³ Steven E. Jacobsen,^{3,4} Wenzhong Wang,¹ Pamela J. Green,^{1,5} and Blake C. Meyers^{1,5}

¹Department of Plant and Soil Sciences & Delaware Biotechnology Institute, University of Delaware, Newark, Delaware 19711, USA; ²Department of Biology, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA; ³Department of Molecular, Cellular and Developmental Biology, University of California, Los Angeles, California 90095, USA; ⁴Howard Hughes Medical Institute, University of California, Los Angeles, California 90095, USA

The *Arabidopsis* genome contains a highly complex and abundant population of small RNAs, and many of the endogenous siRNAs are dependent on RNA-Dependent RNA Polymerase 2 (RDR2) for their biogenesis. By analyzing an *rdr2* loss-of-function mutant using two different parallel sequencing technologies, MPSS and 454, we characterized the complement of miRNAs expressed in *Arabidopsis* inflorescence to considerable depth. Nearly all known miRNAs were enriched in this mutant and we identified 13 new miRNAs, all of which were relatively low abundance and constitute new families. *Trans*-acting siRNAs (ta-siRNAs) were even more highly enriched. Computational and gel blot analyses suggested that the minimal number of miRNAs in *Arabidopsis* is ~155. The size profile of small RNAs in *rdr2* reflected enrichment of 21-nt miRNAs and other classes of siRNAs like ta-siRNAs, and a significant reduction in 24-nt heterochromatic siRNAs. Other classes of small RNAs were found to be *RDR2*-independent, particularly those derived from long inverted repeats and a subset of tandem repeats. The small RNA populations in other *Arabidopsis* small RNA biogenesis mutants were also examined; a *dcl2/3/4* triple mutant showed a similar pattern to *rdr2*, whereas *dcl1-7* and *rdr6* showed reductions in miRNAs and ta-siRNAs consistent with their activities in the biogenesis of these types of small RNAs. Deep sequencing of mutants provides a genetic approach for the dissection and characterization of diverse small RNA populations and the identification of low abundance miRNAs.

[Supplemental material is available online at www.genome.org.]

Nearly all eukaryotes produce small RNAs that function in gene regulation and/or chromatin structure. The small RNA population in plants may be among the most complex because, in addition to producing microRNAs (miRNAs) that play critical roles in various developmental, stress, and signaling responses (for reviews, see Chen 2005; Zhang et al. 2006), plants also produce a complex set of small interfering RNAs (siRNAs) (for review, see Vaucheret 2006). Among the ~77,000 different small RNAs that have been sequenced from *Arabidopsis*, it is likely that miRNAs account for <10%, so the non-redundant set of siRNAs must number >70,000 (Lu et al. 2005). Most of these siRNAs match to repeated sequences such as transposons and retrotransposons. Thus, in cereals and other plant species with larger genomes and correspondingly higher contents of repeated DNA, the complexity of siRNAs is expected to be far greater.

The identification and quantification of small RNAs using deep sequencing methods was first accomplished in *Arabidopsis* (Lu et al. 2005). This revealed, for the first time, the tremendous complexity of plant small RNAs. Specifically, nearly 2.2 million

small RNAs were sequenced from inflorescence and seedlings, yielding >56,000 and 24,000 genome-matching distinct sequences from these tissues, respectively (Lu et al. 2005). Approximately 4300 of these sequences were observed in both libraries. This advance was accomplished using a technology called Massively Parallel Signature Sequencing (MPSS), an unconventional sequencing method that proceeds by a series of cycles involving adapter ligation, hybridization, and enzymatic digestion. The application of MPSS to small RNA sequencing exceeded by more than an order of magnitude the number of small RNAs discovered by all prior conventional sequencing approaches (Llave et al. 2002a; Reinhart et al. 2002; Sunkar and Zhu 2004; Gustafson et al. 2005; Meyers et al. 2006). However, the MPSS analysis was not saturating and limited to two tissue samples, so the full extent of the small RNA population remains to be discovered.

The large number of repeat-associated siRNAs makes it a major challenge to identify low-abundance miRNAs and unusual minor types of siRNAs. Prediction programs have been extremely useful to identify miRNA candidates, yet the output of these programs must be heavily filtered for removal of false positives and this can lead to the loss of bona fide miRNAs (Jones-Rhoades and Bartel 2004; Jones-Rhoades et al. 2006). And although the number of sequenced plant genomes is growing, comparative genomics may also miss important weakly or nonconserved miRNAs. Prediction algorithms also remain to be developed for a particu-

⁵Corresponding authors.

E-mail meyers@dbi.udel.edu; fax (302) 831-4841.

E-mail green@dbi.udel.edu; fax (302) 831-3231.

Article published online before print. Article and publication date are at <http://www.genome.org/cgi/doi/10.1101/gr.5530106>.

larly interesting siRNA type, the *trans*-acting siRNAs (ta-siRNAs), as well as for natural antisense siRNAs, a type thus far observed only under stress conditions (Borsani et al. 2005). Like other siRNAs, the ta-siRNAs require an RNA-dependent RNA polymerase (RdRp) for their synthesis (Peragine et al. 2004; Vazquez et al. 2004; Allen et al. 2005; Yoshikawa et al. 2005). According to current models, production of ta-siRNAs is initiated by miRNA-mediated cleavage of a precursor transcript. The RNA is then reverse-transcribed by RDR6 to form dsRNA, and this dsRNA is processed into 21-nt phased duplexes by a Dicer enzyme called DCL4 (Gascioli et al. 2005; Xie et al. 2005b; Yoshikawa et al. 2005). In the absence of DCL4, DCL2 and DCL3 may also contribute to ta-siRNA production (Gascioli et al. 2005; Xie et al. 2005b; Henderson et al. 2006). Some of the 21-nt species mediate the cleavage of specific mRNAs in *trans*, similar to miRNAs, hence the name ta-siRNAs. Because ta-siRNA production requires an initial miRNA cleavage reaction, the Dicer required for miRNA production, DCL1, is also required for ta-siRNA production (Peragine et al. 2004; Vazquez et al. 2004; Yoshikawa et al. 2005). Neither miRNAs nor ta-siRNAs require *RDR2* for their biosynthesis. Yet, *RDR2* as well as DCL3 are critical for synthesis of the majority of repeat-associated siRNAs, often referred to as heterochromatic siRNAs (Xie et al. 2004). Interestingly, insect and mammal genomes do not appear to encode any *RdRp* genes (Wassenegger and Krczal 2006).

Although the endogenous small RNAs of plants are particularly complex, the availability of mutants in small RNA biogenesis pathways of *Arabidopsis* often makes categorizing small RNAs a straightforward task (for review, see Vaucheret 2006). Typically, an analysis of DCL and RdRp dependencies is performed by preparing RNA gel blots from various mutants and probing for individual small RNAs (Xie et al. 2004). In this report, we sought to take an alternative approach, namely to deeply sequence the *Arabidopsis* small RNA population in several of these small RNA biosynthetic mutants. To this end, we sequenced >900,000 small RNAs from the inflorescence of the *rdr2* mutant, a genotype that is depleted for most repeat-associated, heterochromatic small RNAs. We found the small RNA population from this mutant to be highly enriched for miRNAs, ta-siRNAs, and other interesting *RDR2*-independent siRNAs. By combining deep sequencing, along with computational and gel blot analyses, we were able to identify 13 new miRNAs as well as estimate the minimal number of different miRNA loci in the *Arabidopsis* inflorescence. We also applied 454 sequencing, another parallel sequencing technology (Margulies et al. 2005), to compare small RNA populations of *rdr2* with those of several other mutants of *Arabidopsis*, including *dcl1*, *rdr6*, and a *dcl2/3/4* triple mutant. Taken together, our data indicate that parallel sequencing of different small RNA populations using key mutants is a genetic approach that will greatly contribute to our ability to characterize the small RNAs from plants and other organisms.

Results

Sequencing of *Arabidopsis rdr2* by MPSS and 454

Previous reports have indicated that *rdr2* mutants show a dramatic reduction in endogenous siRNAs and a corresponding increase in miRNAs (Xie et al. 2004). We reasoned that deep sequencing in this mutant would reveal the full complement of miRNAs in *Arabidopsis*. We applied two methods for the high-throughput sequencing of small RNAs (for review, see Meyers et

al. 2006), including the approach we have described based on MPSS (Lu et al. 2005), and a similar approach using the 454 technology (Margulies et al. 2005). MPSS provides extraordinary depth, sequencing a half million or more molecules per library, while 454 has longer reads and thereby provides information about length. Both methods provide quantitative data based on the frequency of the molecules that were sequenced. The small RNA molecules were isolated by size fractionation, sequentially ligated to RNA adapters at the 5' and 3' ends, and used to make cDNA template for sequencing. We generated the libraries using mixed stage inflorescences, which are known to be a rich source of small RNAs (Lu et al. 2005). MPSS produced 915,856 17-nt signatures from *rdr2* (Table 1), which is comparable with the 721,044 signatures previously obtained for wild-type *Arabidopsis* inflorescence (Lu et al. 2005). However, the *rdr2* complexity was reduced by more than 80% compared with wild type in terms of sequence diversity (9066 different genome-matched sequences in *rdr2* compared with 56,920 in wild type). This dramatic difference was despite the larger total number of sequencing reads. Similarly, the 454 sequencing data demonstrated a reduced complexity for *rdr2* small RNAs. Using 454, we sequenced 11,631 small RNAs from wild-type inflorescence (5713 distinct, genome-matching) and 7134 from *rdr2* (686 distinct, genome-matching). The *rdr2* diversity was <13% that of wild type, although in the case of the 454 data, fewer small RNAs were sequenced than with MPSS. The MPSS and 454 data correlated much better for the *rdr2* mutant than the wild type, probably because the reduced complexity of *rdr2* allowed a more saturating level of sampling for even low levels of sequences (Supplemental Fig. S1).

Because *rdr2* is known to lack many heterochromatic siRNAs (Xie et al. 2004), we compared wild-type and *rdr2* sequences to determine if the small RNAs remaining in *rdr2* are primarily a subset of those in wild type. As measured by both MPSS and 454, ~20% of the *rdr2* small RNAs were also observed in the wild-type library (Fig. 1A,B). We hypothesized that this low level of simi-

Table 1. Summary statistics of MPSS and 454 libraries of *rdr2* and wild-type inflorescence

A. MPSS libraries

#	Library	Signatures sequenced ^a	Distinct signatures ^b	Genome matches ^c
1	Wild type (FLR)	721,044	67,528	56,920
2	<i>rdr2</i>	915,856	15,325	9066
Total of #1 and #2		1,636,900	80,741	64,274

B. 454 libraries

#	Library	Signatures sequenced ^a	Distinct signatures ^b	Genome matches ^c
3	Wild type (Col-0)	11,631	9323	5713
4	<i>rdr2</i>	7134	2003	686
Total of #3 and #4		18,765	11,064	6253

^aThe signatures sequenced for each library reflect the sum of two sequencing reactions. "Total" is the sum of the different libraries. Numbers for the 454 data indicate only those sequences for which both 5' and 3' adapters were identified and removed, and the insert was ≥ 15 bp in length.

^b"Distinct" refers to the number of different sequences found within the set. "Total" is the union of the libraries.

^c"Distinct signatures are counted that perfectly match to at least one location in the genome and include signatures matching to tRNAs, rRNAs, snRNAs, or snoRNAs. "Total" is the union of the libraries.

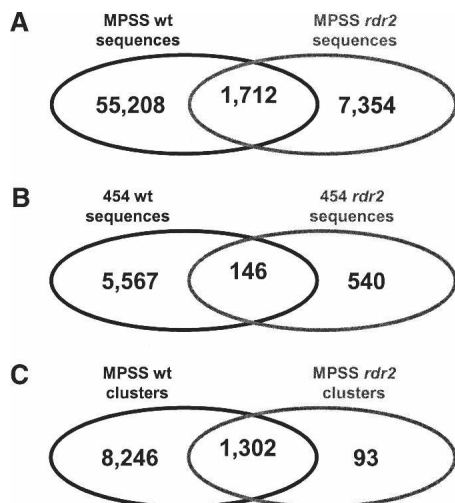


Figure 1. Small RNAs or clusters common to wild type and *rdr2*. Venn diagrams representing genome-matched *rdr2* 454 and MPSS sequences from Table 1. (A) A comparison of distinct signatures in the MPSS libraries indicates 19% of *rdr2* sequences were also found in wild type. (B) A comparison of distinct signatures in the 454 libraries indicates 21% of *rdr2* sequences were also found in wild type. (C) A comparison of genomic clusters of MPSS signatures indicates 93% of small RNA clusters represented in *rdr2* were also found in wild type. For this analysis, clusters contained at least three small RNAs across both libraries; this cutoff was chosen arbitrarily to remove clusters with only one or two small RNAs that could be background. Most of the *rdr2*-only clusters are low-abundance miRNAs or other “real” sequences that were not detected due to depth of coverage in the wild-type library.

larity was largely the result of different siRNAs that represent the same regions. Therefore, we next determined if the genomic loci generating small RNAs in *rdr2* were the same as wild type. To do this, we clustered the small RNAs in both *rdr2* and wild type using our proximity-based algorithm (Lu et al. 2005) and compared clusters across the two libraries for the MPSS data. This analysis demonstrated that nearly all of the clusters (93%) containing at least three small RNAs that were detected in *rdr2* were also detected in the wild-type inflorescence (Fig. 1C). Therefore, most of the small RNA-producing loci in *rdr2* are also producing small RNAs in wild-type inflorescences. Most of the *rdr2*-only clusters were low-abundance sequences that may not have been detected in wild type due to the complexity of wild-type small RNAs and an unsaturated sample size.

We next examined the population of miRNAs in the *rdr2* mutant compared with wild type. The most obvious trend was the expected enrichment of nearly all miRNAs in *rdr2* compared with the wild type (Table 2; Supplemental Table S1). The overall enrichment of miRNAs in *rdr2* was 1.8-fold, based on the proportion of small RNAs represented by known miRNAs (Table 3), a level similar to the 2.2-fold enrichment reported for a low level of sequencing (Xie et al. 2005a). Eight miRNAs were enriched more than fivefold in *rdr2*, including miR158, miR163, miR171, miR172, miR173, miR393, miR399, and miR402 (Table 2). The most abundant miRNA in *rdr2* was miR172. This miRNA was also the most abundant in a *dcl2/3/4* triple mutant (Henderson et al. 2006), which, as discussed below, has a small RNA profile similar to *rdr2*. Both of these mutants lack many common siRNAs, and perhaps this indirectly and positively impacts miR172 abundance. At the other extreme, miR167 had a lower abundance in *rdr2* than wild type, that was most prominent in the MPSS data.

Across the remaining miRNAs, relatively few qualitative differences were observed in terms of miRNAs that were present or absent (Table 2; Supplemental Table S1). For example, the MPSS data showed that only two known miRNA families were present in *rdr2* that had not been detected in wild-type inflorescence (miR157, miR400), while only miR395 was observed in wild-type but not the *rdr2* 454 library (and this may be due to the low sampling depth of the 454 data). Fifteen known miRNAs were never observed in either wild-type or *rdr2* libraries (Table 2; Supplemental Table S1); this could indicate that these miRNAs are not expressed in the tissues or conditions that we sampled, some of these are not bona fide miRNAs as previously suggested (Xie et al. 2005a), or biases in cloning and/or sequencing steps led to their absence.

The *rdr2* small RNAs showed a much more limited distribution on the *Arabidopsis* chromosomes compared with wild type, due to their reduced complexity. The small RNAs from the *rdr2* mutant did not show a pericentromeric concentration, which is a noticeable contrast with wild-type small RNAs (Fig. 2A,B); this is consistent with a loss of heterochromatic siRNAs in *rdr2* (Lippman and Martienssen 2004; Xie et al. 2004). However, there were many more loci matching small RNAs in *rdr2* than are represented by the 117 known miRNA loci (Fig. 2B). This could indicate that many miRNAs, ta-siRNAs, or other *RDR2*-independent small RNAs have yet to be described. As a first step to determine the nature of these *RDR2*-independent small RNAs, we examined the relationship between *rdr2* small RNAs and different genomic regions. Compared with wild type, small RNAs were reduced in *rdr2* in each class of genomic sequence that we investigated (Table 3; Supplemental Fig. S2). Based on the normalized abundances, there was a proportionally greater reduction in small RNAs associated with pseudogenes, transposons, and retrotransposons, compared with genes and unclassified intergenic regions, consistent with a loss of heterochromatic siRNAs (Table 3; Fig. 2B; Supplemental Fig. S2). Small RNAs in the intergenic regions potentially represent unannotated miRNAs or siRNAs from unannotated repeats such as tandem or inverted genomic repeats (Martienssen 2003; Lippman and Martienssen 2004; Lu et al. 2005). Inverted repeats showed one of the lowest reductions in small RNAs in *rdr2*, while small RNAs from tandem repeats were fewer but still well-represented; these two classes of repeats are discussed in more detail below. Because miRNAs are one of the most well-known and interesting classes of small RNAs, and because they do not require *RDR2*, we focused first on the identification of new miRNAs from the *rdr2* sequence data.

Experimental validation of novel miRNAs

As a first step toward the identification of novel miRNAs, *rdr2* MPSS sequences were compared with previously identified wild-type small RNAs in a five-way Venn diagram (Fig. 3). Among those small RNAs that are present in both libraries, the sequences were chosen for further analysis from boxes 3–6 and 9–12; these sequences matched genomic regions that can form hairpin structures and they passed the sparse cluster filter typical of miRNAs (Lu et al. 2005). Eliminating known miRNA genes (101 sequences) and transposons (eight sequences) resulted in a set of 54 small RNA sequences and a total of 31 candidate genomic loci. Because most of the novel candidate miRNAs were sequenced by MPSS multiple times and all were independently detected in two different samples (*rdr2* and wild type), they may represent good

Table 2. miRNA families matched by small RNAs from *rdr2* and wild-type inflorescence

miRNA	MPSS wt (TPQ)	MPSS <i>rdr2</i> (TPQ)	454 wt (raw)	454 <i>rdr2</i> (raw)	454 <i>dcl2/3/4</i> (raw)	454 <i>rdr6</i> (raw)	454 <i>dcl1-7</i> (raw)
<i>miR156</i>	45	976	1	9	1	11	1
<i>miR157</i>	0	684	2	4	4	38	0
<i>miR158</i>	74	3247	3	8	8	71	0
<i>miR159</i>	61	82	246	281	452	398	41
<i>miR160</i>	597	1389	4	5	16	11	2
<i>miR161</i>	913	4248	22	54	73	208	36
<i>miR162</i>	275	918	4	6	21	15	2
<i>miR163</i>	74	15,044	52	210	233	82	0
<i>miR164</i>	467	1560	3	6	11	2	0
<i>miR165</i>	1037	1059	10	25	55	38	5
<i>miR166</i>	10,620	3993	135	174	441	263	16
<i>miR167</i>	59,561	11,061	172	134	331	1270	2
<i>miR168</i>	2267	2091	8	2	37	17	36
<i>miR169</i>	7650	14,488	121	264	20	519	2
<i>miR170</i>	15,704	10,180	52	98	61	122	0
<i>miR171</i>	313	6477	76	89	97	28	10
<i>miR172</i>	1920	93,582	534	2100	1921	329	33
<i>miR173</i>	518	4010	9	44	19	44	0
<i>miR319</i>	372	433	8	25	16	8	0
<i>miR390</i>	11,349	17,445	25	158	84	7	0
<i>miR393</i>	49	972	4	8	10	31	0
<i>miR394</i>	80	382	1	3	3	3	1
<i>miR395</i>	13	23	1	0	0	0	0
<i>miR396</i>	820	1611	9	5	52	28	4
<i>miR397</i>	0	0	0	0	2	0	0
<i>miR398</i>	111	228	1	2	35	18	0
<i>miR399</i>	9	91	0	0	1	7	0
<i>miR400</i>	0	109	0	0	0	16	1
<i>miR401</i>	0	0	0	0	0	0	0
<i>miR402</i>	6	123	0	0	0	0	0
<i>miR403</i>	73	306	2	2	2	4	0
<i>miR404</i>	0	0	0	0	0	0	0
<i>miR405</i>	0	0	0	0	0	0	0
<i>miR406</i>	0	0	0	0	0	0	0
<i>miR407</i>	0	0	0	0	0	0	0
<i>miR408</i>	381	115	1	1	9	1	3
<i>miR413</i>	0	0	0	0	0	0	0
<i>miR414</i>	0	0	0	0	0	0	0
<i>miR415</i>	0	0	0	0	0	0	0
<i>miR416</i>	0	0	0	0	0	0	0
<i>miR417</i>	0	0	0	0	0	0	0
<i>miR418</i>	0	0	0	0	0	0	0
<i>miR419</i>	0	0	0	0	0	0	0
<i>miR420</i>	0	0	0	0	0	0	0
<i>miR426</i>	0	0	0	0	0	0	0
<i>miR447</i>	0	0	0	0	0	0	0
Total from genome ^a			7488	4573	6214	6441	8663

^a“wt” indicates wild type. Values indicate TPQ (MPSS) or raw (454) abundance for perfect matches to known miRNAs with matches located within one nucleotide of the annotated 5' end of the miRNA. Loci with the same name were combined for this analysis; sequences matching individual loci are described in Supplemental Table S1.

^bBecause the 454 values are raw values and not normalized, this row indicates the number of genome-matching small RNAs sequenced in each 454 library as a reference for the miRNA abundance.

candidates for novel *Arabidopsis* miRNAs that are expressed at low levels, may not be conserved between plant species, and have not been described as miRNAs by previous approaches or experiments.

As a complementary experimental approach to validate candidate miRNAs, the expression of candidate miRNAs in different genetic backgrounds was evaluated by RNA gel blot analysis of low molecular weight RNA isolated from inflorescence tissues. Canonical miRNAs generally require *DCL1* (not *DCL2*, 3 or 4), but not *RDR2* or *RDR6*, while 21-nt siRNAs from ta-siRNA loci require *DCL1*, *DCL4*, and *RDR6* but not *RDR2*. *Arabidopsis* mu-

tants with defects in Dicer and RdRp genes, therefore, are important tools to distinguish among different classes of small RNAs (Xie et al. 2004). Of the 31 candidate hairpin-forming genomic loci from the Venn diagram, we conducted RNA gel blot analysis of 13 from boxes containing small RNA signatures with an MPSS abundance of ≥ 40 TPQ, including three small RNAs that we previously predicted to be miRNAs (Lu et al. 2005). Bands within the size range of 21–24 nt expected for mature miRNAs were observed for 12 of 13 candidates that we tested and, of these, nine small RNAs had genetic requirements similar to those of typical, known miRNAs (Fig. 4; Supplemental Table S2A); our blots indicated the small RNAs are present in inflorescence tissue of wild type, *rdr2*, *rdr6*, and a *dcl2/3/4* triple mutant (Henderson et al. 2006), but are absent in *dcl1-7*. Furthermore, these nine small RNAs can form stable foldback structures with the flanking genomic sequence, which is typical of a miRNA precursor, and contain the sequenced small RNA within one arm of the hairpin (Supplemental Fig. S3). Like the majority of known miRNAs, the first 5' nucleotide of these new miRNAs was predominantly a uracil residue (Zhang et al. 2006). Based on the mutant analysis and folding, these are new miRNAs. We focused on boxes 3, 9, and 10 (Fig. 3) to identify new miRNAs because sequences in these boxes lacked a match in AtSet2, indicating that the *Arabidopsis* hairpin sequences were not well conserved with rice. Thus, it is not surprising that among these nine new *Arabidopsis* miRNAs, five do not have identifiable homologs in rice or *Medicago truncatula* based on sequence rather than hairpin comparisons. Like other nonconserved miRNAs, such as miR161 and miR163, these five miRNAs are represented by single loci rather than multigene families (Allen et al. 2004; Jones-Rhoades et al. 2006).

Plant miRNAs function in the regulation of gene expression either by inducing cleavage of their mRNA targets or by translational repression. Therefore, to characterize the function of the new miRNAs we have identified, we predicted the regulatory

targets using an algorithm similar to the one described by Jones-Rhoades and Bartel (2004). In general, cleavage is predominant and can be experimentally assessed using a modified 5'-RACE approach to validate these mRNA targets (Llave et al. 2002b; Reinhart et al. 2002; Kasschau et al. 2003; Jones-Rhoades and Bartel 2004). Targets were predicted with a penalty score of 2.5 or better for seven of the nine new miRNAs (Supplemental Table S3A), using the 21-nt sequence derived from the 17-nt MPSS tag plus four adjacent nucleotides from the matching genomic location. The new *Arabidopsis* miRNA genes we have identified are expressed at relatively low abundances as demonstrated by our

Table 3. Small RNAs from MPSS libraries matching different types of repeats

Type	Wild type		<i>rdr2</i>	
	# distinct signatures	Sum of abundance ^c	# distinct signatures	Sum of abundance ^c
Known miRNA	60	114,732	75	196,194
Known ta-siRNA locus	77	1002	415	22,130
Gene	11,455	135,340	3350	185,367
Pseudogene	1936	8846	53	349
Intergenic regions	30,632	240,505	3583	252,315
Tandem repeats	9423	42,229	1050	18,244
Inverted repeats	3851	24,069	2252	21,688
Retrotransposons ^a	11,533	42,769	189	1905
Transposon ^a	8737	33,198	119	2943
Centromeric ^b	5200	21,615	80	431
rRNA, tRNA, snoRNA, or snRNA	1622	—	258	—

^aNumbers of retrotransposons and transposons include sequences annotated as genes in the TIGR annotation as well as those intergenic regions identified as retrotransposons and transposons by low stringency analysis with RepeatMasker.

^bCentromeric repeats were defined based on regions matching the 180-bp centromeric repeats by BLAST analysis with an E -value $< e^{-10}$.

^c"Sum of abundance" is the sum of TPQ-normalized abundances for all locations of all matching signatures. Signatures with multiple matches in the genome were counted for each type of genomic region in which they matched. Values are not indicated for the type "rRNA, tRNA, snoRNA, or snRNA" because the abundances for these signatures were excluded from our analysis and were not normalized.

MPSS data and RNA gel blots (Fig. 4), and most of them were also absent or marginally represented in other small RNA libraries sequenced by traditional methods (Gustafson et al. 2005). Consequently, mapping of cleavage products generated from these new miRNAs may be challenging due to the low and/or differential expression of the predicted target mRNAs.

Nevertheless, we verified three new miRNA targets among which two have a predicted role in plant defense responses. Two transcripts encoding the CC-NBS-LRR class of putative disease resistance proteins (At5g43740 and At1g51480) were experimentally validated as *in vivo* targets of miR772 (Supplemental Fig. S4A). The predicted target site for miR772 is the region encoding the P-loop domain which is highly conserved in this class of CC-NBS-LRR disease resistance proteins (Meyers et al. 2003). Because of this conservation, miR772 is predicted to target at least 10 more relatives of this gene family (Supplemental Table S3A); the targeting of multiple members of a gene family by a miRNA has previously been reported for several known miRNAs (Llave et al. 2002b; Kasschau et al. 2003; Jones-Rhoades and Bartel 2004). Interestingly, two additional cleavage sites in At1g51480 were mapped, one 31 nt upstream and the other 16 nt downstream of the expected miR772 cleavage site

(data not shown). This may result from the activities of other small RNAs that have not yet been identified. MiR773, miR774, and miR778 were also predicted to target several members of a gene family; for instance, miR774 is predicted to target transcripts for two genes that encode F-box proteins (Supplemental Fig. S4B; Supplemental Table S3A). Notably, several other F-box mRNAs are known targets of miRNA394 and 396 (Jones-Rhoades and Bartel 2004), and target validation assays indicated that the mRNA for another member of this extended gene family (At3g19890) is being cleaved by miR774 (Supplemental Fig. S4B). Although multiple attempts failed to confirm miR778 and miR773-mediated cleavage, the cleavage products of the transcripts predicted to be targets of these miRNAs may be detected in the future, under different conditions that elevate their abundance, for example. These predicted targets include components associated with silencing: two putative SU(VAR)3-9 like histone methyltrans-

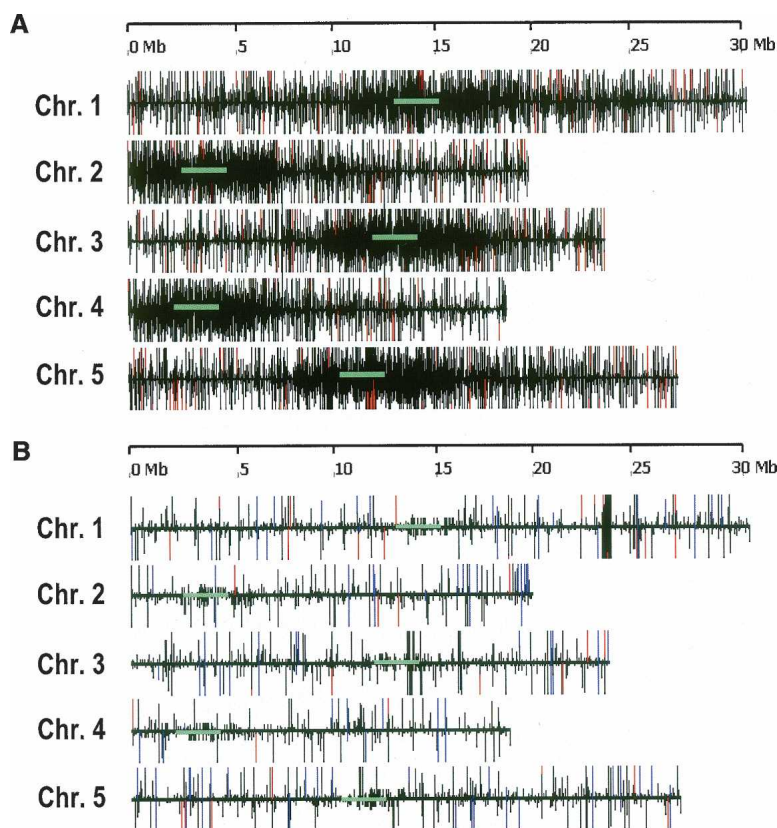


Figure 2. Chromosomal distribution of small RNAs from wild-type and *rdr2* inflorescence. Inflorescence small RNAs matched to the *Arabidopsis* chromosomes as measured by MPSS. The height of the vertical lines indicates the abundance of the small RNA. Maximum height of black bar, >25 . The red bars indicate signatures with >125 transcripts per quarter million (TPQ) and highest black bars indicate signatures with >25 and ≤ 125 TPQ; other black bars indicate signature with ≤ 25 TPQ. Horizontal green bars indicate the approximate location of the centromere. (A) Wild-type inflorescence (adapted from Lu et al. 2005). (B) *rdr2* inflorescence; the abundance of known miRNAs is indicated by blue lines.

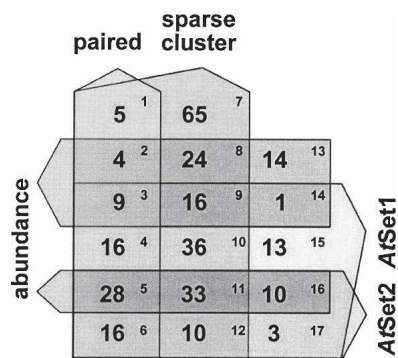


Figure 3. Use of *rdr2* sequences to select miRNA candidates from previously identified wild-type small RNAs. Five-way Venn diagram of selection criteria for miRNAs. The number of distinct *rdr2* MPSS signatures matching the criteria is indicated in each box numbered in upper right; only *rdr2* signatures also found in the wild-type library are represented. The figure excludes 13,153 distinct signatures that did not pass any of the criteria (of which 1583 were found in both *rdr2* and wild-type inflorescence libraries) and 54 matching to the criteria in the Venn which were present in *rdr2* but not wild-type inflorescence (these 54 are included in Supplemental Fig. S2). The paired, sparse, abundance filters, and *AtSet1* and *AtSet2* filters are described elsewhere (Jones-Rhoades and Bartel 2004; Lu et al. 2005) but represent potential hairpin structures typical of miRNA precursors and conservation of those structures in rice, respectively.

ferase (*SUVH5* and *SUVH6*) transcripts that are potential targets of miR778 (Naumann et al. 2005) and members of the family of DNA (cytosine-5)-methyltransferases that are potentially targeted by miR773. Previous reports have described miRNA targets involved in silencing, including *DCL1* and Argonaute1 (*AGO1*), targets of miR162 and miR168, respectively (Rhoades et al. 2002; Xie et al. 2003; Vaucheret et al. 2004).

Estimation of the number of *Arabidopsis* miRNAs in *Arabidopsis*

Given the depth of MPSS coverage and relatively low complexity of *rdr2* small RNA sequences, one could reasonably assume that the majority of inflorescence-expressed *Arabidopsis* miRNAs have been detected in the *rdr2* library (since inflorescence is the richest source of small RNAs among the tissues we have investigated). Using the data from our *rdr2* library, we sought to estimate the total number of miRNAs in *Arabidopsis*. Since we did not want to bias our analysis by including only the best candidates, we reassessed the small RNA sequence data using less stringent criteria. In particular, we allowed into the analysis miRNA candidates only found in the *rdr2* library instead of both libraries (Supplemental Fig. S5) and lowered the acceptable abundance value from 40 TPQ to 10 TPQ. Using this approach, we identified a group of 78 miRNA candidates from the 358 *rdr2* clusters that match to hairpin-like folds and have signatures expressed at levels ≥ 10 TPQ (Fig. 5A). By performing conventional RNA gel blots for 10 candidates that were randomly chosen from the set of 78 and had not already been analyzed during our first round of miRNA validation, we were able to identify an additional two candidate miRNAs (Fig. 5B). Because the RNA gel blot assay for the detection of miRNAs is not particularly sensitive, the expression level of the other eight candidate miRNAs may have fallen below the detection limit of our assay conditions. To increase the sensitivity of detection, we used LNA-modified oligo probes (Valoczi et al. 2004) to test these eight miRNA candidates that were previously undetectable by RNA gel blot analysis. Of these,

four candidates were positive by LNA oligo assay, but the expression of two of the four was not reduced in *dcl1-7* (Fig. 5B). Overall, a total of four of the eight candidates had canonical miRNA expression patterns as detected by either conventional (miR780, miR781) or LNA probes (miR782, miR783) (Fig. 5B). According to miRNA annotation criteria (Ambros et al. 2003), the four of 10 candidates assayed by RNA gel blot hybridization can be considered as bona fide miRNAs (Supplemental Fig. S6); this brings the total number of new miRNAs that we identified to 13. Therefore, these gel blot analyses of the randomly selected small RNAs produced an overall success rate of 40%. Extrapolating from this rate, we would anticipate finding ~31 new miRNAs (from the calculation of $78 \times 40\%$) in the group of 78 miRNA candidates described above, if we had done a comprehensive and exhaustive analysis of inflorescence-expressed small RNAs.

To estimate the total number of *Arabidopsis* miRNAs, two adjustment factors were taken into account for the calculations. Of the 101 known miRNA loci present in the genome-wide folding data of *AtSet1*, 97 were detected by MPSS in the inflorescence libraries. Based on this result, we determined the MPSS detection rate for inflorescence of 0.96 (97 of 101 known miRNAs in *AtSet1*

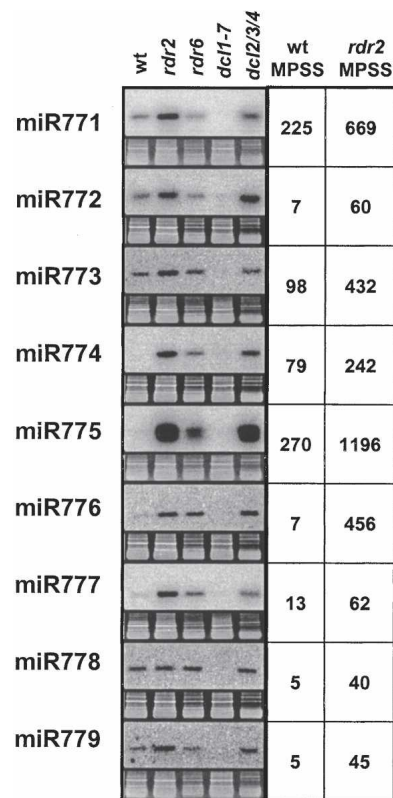


Figure 4. Novel miRNAs identified from Venn analysis of *rdr2* sequences. Small RNAs were selected for validation by RNA gel blots, as described in the text. Low molecular weight RNA isolated from inflorescence tissues was probed with labeled oligonucleotides. The lanes in the blots include the following samples: wild type, *rdr2*, *rdr6*, *dcl1-7*, and *dcl2/3/4*. The normalized abundance level from the MPSS data for *rdr2* and wild type is listed to the right of the identifier for each small RNA. The reason for the apparent increases in abundance in *rdr2* vs. wild type in the blots is not clear; approximately equal amounts of RNA were loaded. It does not appear to be due to RDR2-dependent small RNAs in the 5' flanking regions, although for miR775, the most extreme case, there is an overlapping small RNA that is largely RDR2-dependent that might interfere with miR775 production in wild type.

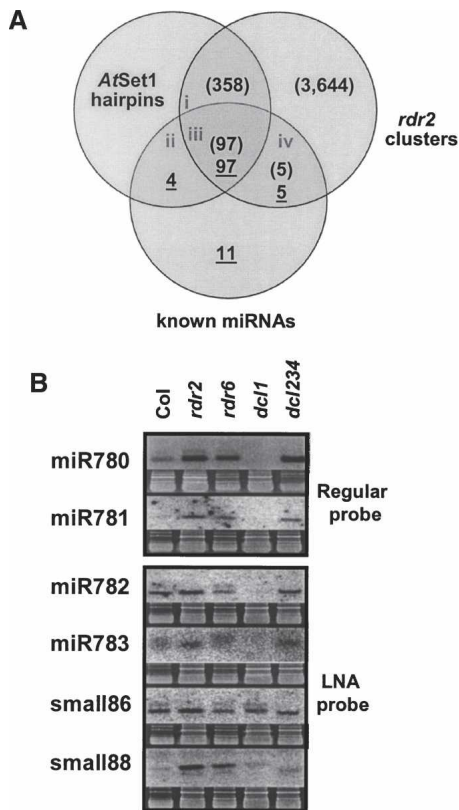


Figure 5. Estimating the minimal *Arabidopsis* miRNA population. (A) A Venn diagram to identify the number of miRNAs that may exist in *Arabidopsis*. Numbers in each section indicate sequences that satisfy each criterion, include *AtSet1* hairpin folds (Jones-Rhoades and Bartel 2004), known miRNAs (underlined numbers), and genomic clusters of *rdr2* MPSS signatures (numbers in brackets). *AtSet1* and *rdr2* clusters were considered matching if the start or end of either overlapped. Lowercase Roman numerals indicate sections from which sequences were selected in B. (B) Gel blots to evaluate the fraction of bona fide miRNAs in selected sectors. Small RNAs identified by *rdr2* signatures in section (i) in A were randomly selected and evaluated by RNA gel blots to determine their presence/absence in the *dcl1-7*, *dcl2/3/4*, and *rdr6* mutants. LNA probes were used for miR782, miR783, small86, and small88 (see Methods).

detected by MPSS). We also determined the *AtSet1* detection rate of 0.86 (101 of 117 known miRNAs in *AtSet1*). Accounting for these two rates, we estimate that there are at least 155 miRNA loci in the *Arabidopsis* genome (calculated from $(97 + 31)/0.96/0.86$, and with the caveat that these sample sizes are too small to draw definitive conclusions). In fact, this is probably an underestimate, if the failure to detect miRNA candidates by our assays was due to very low abundance, restricted cell-type expression, or induction only in response to specific environmental cues (and assuming all annotated miRNAs are real). In addition to the known and new miRNAs, the *rdr2* data demonstrated the presence of numerous RDR2-independent small RNAs, some of which may yet be characterized as new miRNAs.

Other RDR2-independent small RNAs in *Arabidopsis*

We and other groups have previously shown that a significant number of *Arabidopsis* endogenous siRNAs match to various kinds of repeats (Martienssen 2003; Lippman and Martienssen 2004; Lu et al. 2005). Xie et al. (2004) have shown the requirement of

RDR2 and DCL3 for the biosynthesis of a subset of repeat-associated siRNAs. However, considering the presence of multiple RdRps in *Arabidopsis* and the diversity of repeats, it is unclear which populations of siRNAs generated from repeat sequences are dependent on RDR2 activity. We separately characterized the RDR2-dependent and RDR2-independent inverted and tandem repeats, as these repeats are known to be sources of small RNAs. The RDR2-dependent inverted repeat set, comprising a total of 461 genomic locations, was defined as those for which (1) the sum of abundance is ≥ 10 TPQ in wild type; (2) the sum of abundance is at least 10-fold higher in wild type than in *rdr2*. Similarly, we considered a repeat to be RDR2-independent only if the sum of abundance from the repeat is ≥ 10 TPQ and not down-regulated ($rdr2/wt \geq 1$) in *rdr2*; we found 55 loci for this set (12% of the total). As shown in Table 4, the repeat score of the RDR2-independent set was significantly higher than that of the RDR2-dependent set (Mann-Whitney Test: P -value = 0.0048). One of the primary determinants of the score is the length of the repeat, suggesting that the RDR2 dependence of inverted repeats may be based on their length. This is consistent with a previous study suggesting that, for some inverted repeats, RDR2 may contribute to the formation or stability of a complex that contains active DCL3 (Xie et al. 2004). For genomic loci that contain long inverted duplications and can form extensive dsRNA structures (foldbacks), RDR2 is most likely dispensable for siRNA production (Table 4). One hypothesis is that one or more Dicers can efficiently process long dsRNA precursors even in the absence of RDR2. In agreement with this, closer examination of some RDR2-independent inverted repeats revealed that these loci usually showed complex patterns of siRNA accumulation with different size classes affected by different Dicer mutants (Supplemental Fig. S7).

We identified a potential foldback structure in the S-receptor kinase gene (*SRK*) as one of the most strongly expressed RDR2-independent siRNA-producing regions (Supplemental Fig. S8). The large number of sequenced small RNAs matching to this stem-loop suggests that it is a substrate for Dicer cleavage. The observation that small85 from this locus is still evident in the *dcl1-7* and *dcl2/3/4* mutants but not in a quadruple *dcl1/2/3/4* mutant (data not shown) suggests the involvement of multiple Dicers (Supplemental Fig. S8). Functional copies of *SRK* and a gene called *SCR* are important for self-incompatibility in Brassica and *Arabidopsis* species (such as *A. lyrata*) (Nasrallah et al. 2002). Loss of this self-incompatibility system in *Arabidopsis thaliana* is one of the key factors that led to the selection of *A. thaliana* as a model system for plants (Redei 1975). Suggested explanations for this loss include the fragmented *SCR* gene or the alternatively spliced *SRK* transcripts that contain premature nonsense codons that are present in *A. thaliana* (Kusaba et al. 2001). Our data suggest that the *SRK* gene may be silenced by an inverted repeat, and these small RNAs may have played a previously unknown role in the loss of SRK function in *A. thaliana*.

Unlike inverted repeats from which dsRNA is readily generated simply by folding of a single RNA, tandem repeats should require an RdRp to form dsRNA structures. Indeed, tandem repeats show a higher overall dependence on RDR2 than inverted repeats (Table 4). Our RDR2-dependent tandem repeat set contained 3491 genomic locations whereas the RDR2-independent tandem repeat set contained only 82 loci (2% of the total). Interestingly, we also found that the average length of the tandem repeat unit in the RDR2-dependent set is significantly larger than that of the RDR2-independent set (Mann-Whitney Test: P -

Table 4. RDR2-dependent and RDR2-independent repeats from MPSS libraries

A. Inverted repeats				
	Score	% of similarity	Gap ^a	Size
RDR2-dependent	799.4 ± 34.0	86.4 ± 0.45	5.7 ± 0.4	405 ± 17
RDR2-independent	1595.7 ± 232.7	86.7 ± 1.5	7.1 ± 1.1	713 ± 86
B. Tandem repeats				
	Score	% of similarity	Count ^b	Size ^c
RDR2-dependent	129.8 ± 8.1	84.1 ± 0.1	3.7 ± 0.06	101.4 ± 3.4
RDR2-independent	44.1 ± 12.0	81.6 ± 0.9	5.1 ± 0.73	32.8 ± 4.3

In each case, RDR2-dependent is defined as the sum of abundance is ≥ 10 TPQ in wild type and the sum of abundance is at least 10-fold higher in wild type than in *rdr2*; RDR2-independent is defined as the sum of abundance from the repeat is ≥ 10 TPQ in *rdr2* and the small RNAs are not down-regulated in *rdr2* ($rdr2/wt \geq 1$). Mean values for each category are indicated followed by standard error (\pm). The score was determined by the programs Einverted or Etandem and represents a composite of length and identity for each set of repeats. The complete set of inverted and tandem repeat data is provided in Supplemental File 1.

^a“Gap” indicates the average gap between arms of the inverted repeat (in nucleotides).

^b“Count” refers to the number of tandem repeats.

^c“Size” indicates the average length of the repeats at each locus (in nucleotides).

value = 0.0001). Therefore, high quality and long tandem repeats generally appear to require RDR2 to generate dsRNAs and sustain siRNA production. Other RdRps probably facilitate dsRNA production from these short tandem repeats because the *Arabidopsis* genome contains six RdRp homologs (Yu et al. 2003). One likely hypothesis is that different RdRps could function redundantly on tandem repeats.

Known ta-siRNA loci were the most enriched small RNA sources in the *rdr2* background. For the four previously characterized ta-siRNA loci, the sum of small RNA abundance was at least 20-fold higher in *rdr2* than in wild type based on the MPSS data (Table 5A; Supplemental Fig. S9). This greatly exceeds the 1.8-fold for enrichment of total miRNA abundance mentioned earlier. Using known ta-siRNAs as reference, we developed a set of filters to enrich for new ta-siRNAs. Four filters were designed and applied to identify genomic locations representing potential ta-siRNA loci: (1) The cluster contains at least 10 distinct signatures; (2) the sum of abundance for the cluster is ≥ 100 TPQ; (3) the sum of abundance is at least 10-fold higher in *rdr2* than in wild type; (4) the cluster does not match to known miRNAs, ta-siRNAs, transposons, retrotransposons, or centromere repeats. These filters generated 28 potential ta-siRNA loci (Supplemental Table S4). Interestingly, among these, 14 loci (50% of the filter output) corresponded to different members of the PPR gene family, a group of genes known to be targeted by miRNAs, ta-siRNAs, and siRNAs (Allen et al. 2005). Seven of the 14 remaining candidate loci were further examined by RNA gel blotting. We found two candidates (small49, small58) displaying typical ta-siRNA expression patterns (present in *rdr2* but very low in *rdr6*, *dcl1*, and *dcl2/3/4*) (Supplemental Table S2B; Fig. 6). Furthermore, a clear 21-nt phased pattern was observed at the locus containing small49, consistent with Dicer

activity (Fig. 6). With this low stringency filtering protocol that captures all known ta-siRNA loci, we found relatively few loci which had ta-siRNA characteristics. Therefore, we interpret these data as an indication that ta-siRNA genes are rare in the *Arabidopsis* genome. This result is consistent with the observation that mutations that block ta-siRNA production have a relatively weak phenotype (Peragine et al. 2004; Vazquez et al. 2004). However, it is also possible that other ta-siRNAs were expressed at very low levels or not at all under our sampling conditions.

Small RNA size distribution in *rdr2* and the small RNA populations in other mutants

The enrichment of miRNAs and loss of heterochromatic siRNAs in *rdr2* should correlate with a shift in the sizes of the small RNA population. Canonical miRNAs are 21 nt while canonical heterochromatic siRNAs are 24 nt. Because the MPSS sequence data are limited to 17 nt for small RNAs, we used the 454 sequence data to determine the size distribution of the small RNAs. As an additional comparison to wild-type and *rdr2* inflorescences, we also sequenced small RNAs from the inflorescence of the *Arabidopsis* mutants *rdr6* and *dcl1-7*, and compared these with data we recently obtained for *dcl2/3/4* (Henderson et al. 2006). All of these mutants are altered in important genes for small RNA biogenesis. We assessed the size distribution based on both distinct sequences and total abundances (Fig. 7). Both *rdr2* and the *dcl2/3/4* triple mutant showed a similar pattern of 24-nt siRNA reduction and 21-nt miRNA enrichment (Fig. 7A,B). The increase in 21 mers in both mutants reflects an enrichment of miRNAs and is consistent with previous reports (Table 2; Fig. 7B,D; Xie et al. 2004). In contrast to miRNAs, 21-nt siRNAs from known ta-siRNA loci can be readily identified from *rdr2* but were absent in *dcl2/3/4* (Table 5B; Henderson et al.

Table 5. Representation of known ta-siRNA loci in small RNA libraries

A. MPSS libraries						
ta-siRNA locus	Chromosome	Start coordinates (bp)	End coordinates (bp)	# distinct signatures ^a	Sum of abundance in wild type	Sum of abundance in <i>rdr2</i>
TAS1a	2	11,728,344	11,729,168	94	115	7633
TAS1b	1	18,552,926	18,553,725	63	217	13,115
TAS1c	2	16,544,582	16,545,150	126	349	10,456
TAS2	2	16,546,598	16,547,391	92	457	8027
TAS3	3	5,862,059	5,862,369	81	66	2094
B. 454 libraries						
ta-siRNA locus	Wild type (Col-0)	<i>rdr2</i>	<i>rdr6</i>	<i>dcl1-7</i>	<i>dcl2/3/4</i>	
TAS1a	12	41	4	0	0	
TAS1b	7	37	1	0	0	
TAS1c	32	72	3	0	0	
TAS2	28	71	1	0	0	
TAS3	13	11	1	5	0	

^aThe number of distinct signatures was calculated as the sum of distinct signatures in the wild-type and *rdr2* libraries.

2006), consistent with previous observations that DCL4 is required for ta-siRNA production (Gascoilli et al. 2005; Xie et al. 2005b; Yoshikawa et al. 2005; Henderson et al. 2006). Nevertheless, a strong correlation between the 454 data of *rdr2* and *dcl2/3/4* was observed ($R^2 = 0.92$ for all small RNAs present in both libraries; $R^2 = 0.95$ for miRNAs; Supplemental Fig. S10). In contrast, the *dcl1-7* mutant demonstrated a lower proportion of 21-nt small RNAs compared with wild type (Fig. 7B), and most of this difference can be attributed to a substantial reduction in known miRNAs (Table 2; Fig. 7B,D). This is consistent with the known reduction in the miRNA complement of *dcl1-7* (Park et al. 2002). Both the wild type and *rdr6* mutant have substantial peaks at both 21 and 24 nt, as expected. However, analysis of ta-siRNA abundance in the *rdr6* mutants has revealed that indeed very few ta-siRNAs were detected in the absence of RDR6 (Table 5B).

Even the modest depth of the 454 sequencing was sufficient to identify differential effects of specific mutants on the accumulation on miRNA families. Although DCL1 appears to be the only Dicer protein responsible for miRNA biogenesis in *Arabidopsis* (Henderson et al. 2006; Jones-Rhoades et al. 2006), some miRNAs are affected less than others by the *dcl1-7* mutant. The most extreme case was miR168 which did not decrease at all in *dcl1-7*

based on our 454 data (Table 2). These results are in agreement with Vaucheret et al. (2006), who reported no decrease in miR168 levels in three different *dcl1* partial loss-of-function mutants. This fits well with the model that miR168 levels are not limited by DCL1 activity but are instead controlled by a feedback loop involving AGO1, the target of miR168; AGO1 is hypothesized to both stabilize miR168 and also slice its own mRNA using miR168 as a guide (Vaucheret et al. 2006). The accumulation of miR159 and miR165/166 has also been reported to be somewhat less sensitive to *dcl* mutations than other miRNAs tested (Park et al. 2002; Reinhart et al. 2002; Kasschau et al. 2003; Kurihara et al. 2006; Vaucheret et al. 2006) and we also observed these subtleties. Finally, members of the miR161 family and miR408 are known to be rather insensitive to the *dcl1-7* allele (Allen et al. 2004) and the *dcl1-9* allele (Sunkar and Zhu 2004) respectively, results quite consistent with our 454 data. Based on the close recapitulation of published observations with our *dcl1* data, it seems likely that other differential accumulation characteristics resulting from our data set represent regulatory characteristics of potential biological significance. These would include miR167, which is down-regulated in *rdr2* compared with wild type, and miR172 which is of particularly high abundance in *rdr2* and *dcl2/3/4* (Table 2). Another miRNA with unusual characteristics is miR169. This miRNA is an outlier in the correlation of *rdr2* and *dcl2/3/4* (Supplemental Fig. S10), having a very low accumulation in *dcl2/3/4*, with high accumulation in *rdr2*. Given that miR169 is also increased in *rdr6* and encoded by a tandem array of genes, these accumulation results may be due to a secondary level of control by an siRNA-mediated pathway.

Discussion

We have characterized the small RNAs in the *Arabidopsis rdr2* mutant and demonstrated that this mutant contains a dramatically altered population of small RNAs. Many fewer heterochromatic siRNAs were observed, and miRNAs and other specialized categories of small RNAs were enriched, facilitating their identification and classification. We applied two complementary sequencing methods, MPSS, which is very deep, and 454, which provides size information; these combined data sets confirmed and greatly extended earlier studies by providing quantitative, genome-wide small RNA data. Endogenous siRNAs corresponding to transposons, retrotransposons, centromere repeats, and pseudogenes virtually disappear in the *rdr2* background (Fig. 2; Supplemental Fig. S2), as predicted by prior low-depth analyses (Xie et al. 2004; May et al. 2005). A surprising finding was that some of the small RNAs remaining in *rdr2* had significantly higher abundance than in wild type (Supplemental File 1) and derived from repeats that had special characteristics. Other RDR2-independent siRNAs were phased by 21 nt and/or had other features characteristic of ta-siRNAs. Indeed, we found that small RNAs matching to known ta-siRNA loci were dramatically enriched in *rdr2* (5.5% by abundance compared with 0.2% in wild type) to an extent even greater than that of miRNAs (51.5% in *rdr2* compared with 28.9% in wild type) (Table 3). The enrichment of miRNAs in *rdr2* was a tremendous advantage for the identification of new low abundance miRNAs. And finally, the parallel sequencing by 454 of additional libraries from small RNA biosynthetic mutants not only recapitulated published data but also provided new and interesting examples of differential effects on specific miRNAs and siRNAs.

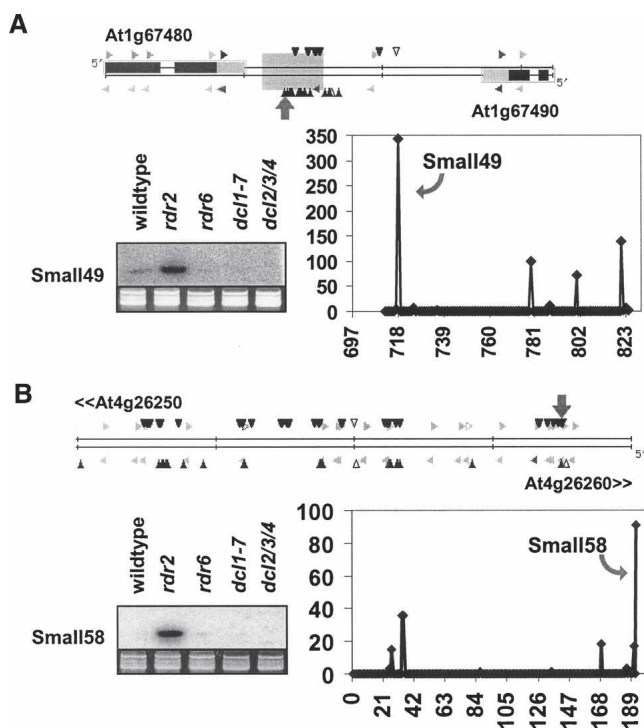


Figure 6. RDR2-independent small RNAs from regions with ta-siRNA-like features. (A) The locus that includes small49 exhibits 21-nt phasing and accumulation characteristics in mutants similar to those of ta-siRNAs. Image of the MPSS Web viewer for the intergenic region that contains small49 in the position indicated by the large gray arrow (the shaded square covering both strands in this intergenic region indicates an inverted repeat); an RNA gel blot of small49; a plot of the y-axis indicating the small RNA abundance in the *rdr2* mutant as measured by MPSS (in TPQ), and the x-axis indicating nucleotide position on Chr. 1, with the "697" indicating position 25,282,697. (B) The blot shows that small58 also has ta-siRNA-like accumulation features; images as described in A, with the 0 position in the x-axis of the plot indicating nucleotide 13,295,900 on Chr. 4. Due to limited space, only a portion of this intergenic region is shown, with the flanking genes indicated at the ends by their At identifiers.

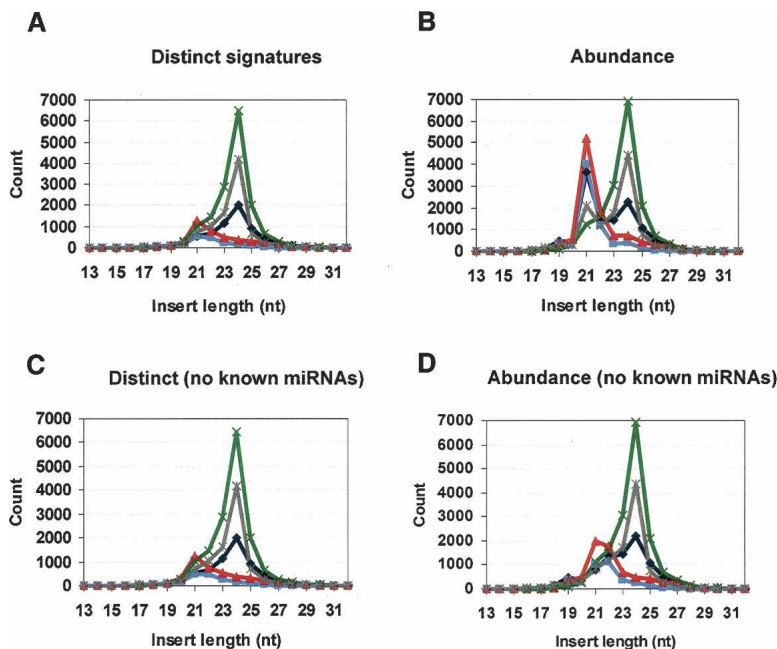


Figure 7. Small RNA size distribution in mutants evaluated with 454 sequencing. In each plot, gray indicates wild type, light blue is *rdr2*, green is *dcl1–7*, dark blue is *rdr6*, and red is the *dcl2/3/4* triple mutant. (A) Number of distinct signatures vs. size. (B) Total abundance of sequences vs. size. (C) Number of distinct vs. size, with known miRNAs removed. (D) Total abundance vs. size, with known miRNAs removed.

miRNA discovery

While most *Arabidopsis* miRNAs have been identified by traditional cloning and sequencing of small RNAs, it is unlikely that these screens are saturating for rare or tissue-specific miRNAs. The need for additional methods of miRNA identification led to the development of bioinformatics methods for the prediction of miRNAs (Bonnet et al. 2004; Jones-Rhoades and Bartel 2004; Wang et al. 2004). Most of these computer algorithms rely on evolutionary conservation of miRNA sequences between different species and, therefore, are limited to the detection of only conserved miRNAs, although at least one analysis has relied only on intra-genomic comparisons (Arai et al. 2005). Even these predictions ultimately require either high-throughput or highly sensitive methods for validation. With MPSS and other high-throughput sequencing technologies, the sequencing of small RNAs is no longer a limiting factor in the discovery of novel miRNAs. However, we have demonstrated that by combining these approaches with mutants in which miRNAs are significantly enriched compared with wild type, such as *rdr2* and *dcl2/3/4*, we can efficiently delineate the small RNAs as miRNAs, siRNAs, or other categories. Even at relatively low sampling depths, many known miRNAs were observed and their abundance was measured. Compared with wild type, the MPSS data for *rdr2* were dramatically simplified and “cleaned up” of siRNAs, making miRNA candidates much easier to identify. Our 454 analysis indicated that the *rdr2* and *dcl2/3/4* triple mutants are most similar in their small RNA profiles, consistent with the idea that these genes may be in the same pathway involved in heterochromatic siRNA production (for review, see Brodersen and Voinnet 2006) and a mutant of either type (*rdr2* and *dcl2/3/4*) enriches for miRNAs.

Prior experimental and computational efforts over the last

several years have resulted in the identification of 117 miRNA genes in *Arabidopsis* which can be grouped into 42 families. The 13 new miRNAs that we report all represent new families that presumably escaped previous discovery because of their low abundance. These new miRNAs increase the total number of *Arabidopsis* miRNA families by 25%. Eight of the newly described miRNAs are found only in *Arabidopsis*. For nonconserved miRNAs, it is more difficult to confidently predict targets because the conservation of the target site cannot be used as a filter to remove false positives. Therefore, a highly stringent score (≤ 2.5) was applied in our target prediction. Potential regulatory targets were found for 10 of the 13 miRNAs. Some of the biological roles of the newly confirmed or predicted targets resemble those of previously identified *Arabidopsis* miRNAs. At least three of these are bona fide because we could map the cleavage products and we predict that others were simply beneath our threshold of detection. MiR774 targets the mRNA of at least one F-box protein. Combined with six previously identified F-box genes, there are at least seven F-box mRNAs targeted by miRNAs, suggesting that the protein degradation machinery is subject to considerable miRNA regulation (Jones-Rhoades et al. 2006). Our observation that miR773 mediates the cleavage of at least two, and potentially more, members of the CC-NBS-LRR class of putative disease resistance proteins suggests a previously unknown role of miRNAs in plant defense. As new and more sensitive methods for verifying miRNA targets are developed, it will be exciting to see if some of the other interesting putative targets such as the methyltransferases in Supplemental Figure S4 can be verified. While, our target predictions focused on protein-coding genes, at least two miRNAs (mir173 and mir390) target precursors of ta-siRNAs; consequently, there may be additional targets for some of these new miRNAs that have not yet been identified.

From our analyses, we have predicted based on our sequence data and RNA gel blot assays that at least 155 miRNAs should exist in the *Arabidopsis* genome. We believe that this is a conservative estimate. Several groups have shown that RNA gel blots are not sensitive enough to detect many real miRNAs (Grad et al. 2003; Bentwich et al. 2005; Grundhoff et al. 2006). Moreover, if a higher percentage of the candidate miRNAs that were negative by gel blot analysis can be detected by more sensitive detection methods such as PCR, or if we loosen our abundance cutoff to below 10 TPQ, more miRNAs may appear to be bona fide. It should be noted also that inflorescence tissues were used for many of the previous miRNA cloning projects, and, as a result, the set of known miRNAs likely is biased toward those that are expressed in flowers. Therefore, the real number of miRNAs may well exceed our estimate. Interestingly, many of the new miRNAs identified in this study did not have homologs in rice, suggesting that conservation between *Arabidopsis* and rice may not be one of the best filters for identifying additional new miRNAs. The con-

ervation of miRNA families across plant subgroups such as eudicots, monocots, gymnosperms, and even moss has been reported (Arazi et al. 2005; Axtell and Bartel 2005; Zhang et al. 2006). However, the presence of eight *Arabidopsis*-specific miRNAs indicates substantial miRNA evolution in the time since the divergence of monocots and dicots, and suggests that miRNA enrichment methods such as our use of the *rdr2* mutant may be valuable in many plant species.

RDR2-independent siRNAs

Tandem repeats are prone to epigenetic silencing mediated by RNA interference (Martienssen 2003). Previous studies have shown that several siRNAs corresponding to tandem repeats in the *Arabidopsis* genome were absent in *rdr2* (Xie et al. 2004; May et al. 2005). It has been proposed that tandem repeats can sustain RdRp activity because the first round siRNAs can randomly initiate subsequent rounds of siRNA production and perpetuate the siRNA pool (Martienssen 2003). While this model has not been proven, it is substantiated by our MPSS data indicating that almost all the tandem repeats in the *Arabidopsis* genome required RDR2 activity to generate siRNAs. However, for some of these tandem repeats, the small RNAs were significantly higher in *rdr2* than in wild type (Supplemental File 1). Something about these tandem repeats, perhaps their relatively low quality, may allow these sequences to be silenced independently of RDR2. In this case, other components of the siRNA biogenesis machinery must be involved in the recognition and generation of siRNAs from these specific loci. This suggests that the biogenesis pathway for repeat-associated siRNAs is more complex than initially believed and the production of some repeat-associated siRNAs does not require RDR2 activity.

Our understanding of the mechanism by which siRNAs are derived from inverted repeats is less clear. siRNA accumulation from inverted-repeat loci is dependent on RDR2 and DCL3 (Xie et al. 2004). While DCL3 clearly functions as the ribonuclease to process dsRNA precursors, it is unclear why RDR2 is essential to this pathway. Another example is siRNA production from constructs used for inverted-repeat post-transcriptional gene silencing (IR-PTGS, typically used for RNAi). Although widely used as a research tool, IR-PTGS remains one of the least understood plant RNA silencing processes. Until recently, no mutant defective in this pathway had been recovered, and IR-transgene-induced siRNA accumulation is not affected by single-gene mutations (for review, see Brodersen and Voinnet 2006). Our analysis of *rdr2* by MPSS may provide an explanation for these apparently contradictory observations. In agreement with previous studies, the majority of endogenous inverted repeats, such as the siRNA02 locus (Xie et al. 2004), did not accumulate siRNAs in the absence of RDR2. However, we also identified a group of inverted repeats which produced siRNAs independently of RDR2 (Supplemental File 1). One difference between RDR2-dependent and RDR2-independent inverted repeats is that the latter set tends to have a higher repeat score and larger size of repeat unit. Although it is difficult to rule out alternative hypotheses completely, the simplest interpretation of the data is that RDR2 and DCL3 are required for only a subset of inverted repeats, generally with low scores and relatively short repeat units. In the case of longer and higher scoring inverted repeats, RDR2 activity (and probably DCL3) may not be required, similar to IR-transgenes. One likely scenario is that the high quality dsRNA structures generated from long inverted repeats are subject to the activity of different Dicers

(Brodersen and Voinnet 2006). Consistent with this model, recent analyses of combinatorial Dicer knockout mutants indicated that the functions of different *Arabidopsis* Dicer proteins are highly redundant (Gascioli et al. 2005; Henderson et al. 2006).

Conclusions

Clearly, analyses of plant miRNAs may achieve a higher level of sensitivity by using an *rdr2* background. This approach may prove even more beneficial in other plant species, particularly those with larger and more repeat-laden genomes. The combined deep profiling data from MPSS and full-length sequencing of small RNAs from different genotypes by 454 demonstrate that small RNA sequence libraries are a rich and novel source of data that have yet to be fully exploited in *Arabidopsis* or any other organism. As sequencing costs drop with the advent of new short-read sequencing technologies, the approaches that we have implemented for the analysis of *Arabidopsis* mutants are likely to be more broadly applied for experimental investigation of different conditions or mutants.

Methods

Plant growth

All plant material was from *Arabidopsis* ecotype Col-0. The *rdr2*, *rdr6*, *dcl1-7*, and *dcl2/3/4* mutants have been described previously (Peragine et al. 2004; Xie et al. 2004; Henderson et al. 2006). Inflorescence tissue was harvested from plants grown in soil in a growth chamber with 16 h of light for 5 wk. Floral tissue included the inflorescence meristem and early stage floral buds (up to Stage 11/12). Total RNA was isolated using Trizol reagents (Invitrogen). Inflorescence and seedling material were harvested approximately at eight hours into the subjective day.

RNA gel blot analysis

Total RNA was extracted with Trizol (Invitrogen). Low molecular weight (LMW) RNA gel blots, radiolabeled probes for specific small RNAs, and hybridization/wash conditions were described previously (Lu et al. 2005). Blot hybridization analysis was performed as described (Llave et al. 2002a). All blots shown are representative of at least two independent experiments.

Locked nucleic acid (LNA) probes (Valoczi et al. 2004) were used as indicated in the figure legends; these probes were used when the hybridization signal was not detectable using regular oligonucleotides. LNA oligos were obtained from Sigma-Proligo. Hybridization conditions were as described (Valoczi et al. 2004).

MPSS and 454 data generation and analysis

All MPSS sequencing and analysis were performed essentially as described (Brenner et al. 2000a,b; Meyers et al. 2004; Lu et al. 2005). The small RNA libraries were constructed as previously described (Lu et al. 2005; Meyers et al. 2006). The raw and normalized MPSS data are available at <http://mpss.udel.edu/at>. 454 analysis was performed essentially as described (Margulies et al. 2005). Adapter sequences were identified and removed using local alignments. The summary statistics of the *rdr2* and wild-type 454 libraries are described in the text; the *dcl1-7* and *rdr6* libraries included 12,060 and 16,856 adapter-trimmed small RNA inserts, respectively, and the *dcl2/3/4* triple mutant 454 library has recently been described (Henderson et al. 2006).

We compared MPSS signatures with the TIGR annotation version 5.0 (Wortman et al. 2003) and assigned signatures to each location at which a perfect match was found. The number

of matches was recorded as the “hits.” As previously described (Lu et al. 2005), we merged the MPSS sequencing runs and calculated a single abundance normalized to “transcripts per quarter million” (TPQ) after the removal of rRNAs, tRNAs, snoRNAs, or snRNAs signatures. Clustering of small RNAs was based on the previously described proximity-based algorithm, with the same setting of a 500-bp window for the clusters that was used in our prior analysis (Lu et al. 2005). Repeat analysis was also performed as described previously (Lu et al. 2005) using a combination of programs including RepeatMasker (<http://www.repeatmasker.org/>), Einverted and Etandem (Rice et al. 2000).

Accession numbers

Sequence data from this article can be found in the Gene Expression Omnibus database at NCBI (<http://www.ncbi.nlm.nih.gov/geo/>). The accession numbers for GEO are as follows: Series identifier GSE5343, platform identifier GPL4011, and sample identifiers are GSM120717 (*rdr2* MPSS data), and GSM121453 to GSM121457 (454 data). The raw and normalized MPSS data are also available at <http://mpss.udel.edu/at>, and this Web site allows users to query these data based on physical location, gene identifiers, or by sequence.

Acknowledgments

We thank Mayumi Nakano for her work on the Web interface, Kalyan Vemaraju for his work on the MPSS database, Deborah Charlesworth for discussions of the *SRK* locus, and several anonymous reviewers for their valuable comments. This work was supported primarily by NSF grants 0439186 and 0548569 (P.J.G. and B.C.M.), with additional support and resources provided by NSF Plant Genome grant 0321437 (B.C.M.), DOE DE-FG02-04ER15541 (P.J.G.), and NIH P20 RR16472-04. I.R.H. was supported by an EMBO Long-term Fellowship. S.E.J. is a Howard Hughes Medical Institute Investigator, and work in S.E.J.’s laboratory was supported by NIH grant GM60398. Research in the Poethig laboratory was supported by NIH grant GM51893.

References

- Aday, A., Johnson, C., Mlotshwa, S., Archer-Evans, S., Manocha, V., Vance, V., and Sundaresan, V. 2005. Computational prediction of miRNAs in *Arabidopsis thaliana*. *Genome Res.* **15**: 78–91.
- Allen, E., Xie, Z., Gustafson, A.M., Sung, G.H., Spatafora, J.W., and Carrington, J.C. 2004. Evolution of microRNA genes by inverted duplication of target gene sequences in *Arabidopsis thaliana*. *Nat. Genet.* **36**: 1282–1290.
- Allen, E., Xie, Z., Gustafson, A.M., and Carrington, J.C. 2005. microRNA-directed phasing during *trans*-acting siRNA biogenesis in plants. *Cell* **121**: 207–221.
- Ambros, V., Bartel, B., Bartel, D.P., Burge, C.B., Carrington, J.C., Chen, X., Dreyfuss, G., Eddy, S.R., Griffiths-Jones, S., Marshall, M., et al. 2003. A uniform system for microRNA annotation. *RNA* **9**: 277–279.
- Arazi, T., Talmor-Neiman, M., Stav, R., Riese, M., Huijser, P., and Baulcombe, D.C. 2005. Cloning and characterization of micro-RNAs from moss. *Plant J.* **43**: 837–848.
- Axtell, M.J. and Bartel, D.P. 2005. Antiquity of microRNAs and their targets in land plants. *Plant Cell* **17**: 1658–1673.
- Bentwich, I., Avniel, A., Karov, Y., Aharonov, R., Gilad, S., Barad, O., Barzilai, A., Einat, P., Einav, U., Meiri, E., et al. 2005. Identification of hundreds of conserved and nonconserved human microRNAs. *Nat. Genet.* **37**: 766–770.
- Bonnet, E., Wuyts, J., Rouze, P., and Van de Peer, Y. 2004. Detection of 91 potential conserved plant microRNAs in *Arabidopsis thaliana* and *Oryza sativa* identifies important target genes. *Proc. Natl. Acad. Sci.* **101**: 11511–11516.
- Borsani, O., Zhu, J., Verslues, P.E., Sunkar, R., and Zhu, J.K. 2005. Endogenous siRNAs derived from a pair of natural *cis*-antisense transcripts regulate salt tolerance in *Arabidopsis*. *Cell* **123**: 1279–1291.
- Brenner, S., Johnson, M., Bridgham, J., Golda, G., Lloyd, D.H., Johnson, D., Luo, S., McCurdy, S., Foy, M., Ewan, M., et al. 2000a. Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. *Nat. Biotechnol.* **18**: 630–634.
- Brenner, S., Williams, S.R., Vermaas, E.H., Storck, T., Moon, K., McCollum, C., Mao, J.I., Luo, S., Kirchner, J.J., Eletr, S., et al. 2000b. In vitro cloning of complex mixtures of DNA on microbeads: Physical separation of differentially expressed cDNAs. *Proc. Natl. Acad. Sci.* **97**: 1665–1670.
- Brodersen, P. and Voinnet, O. 2006. The diversity of RNA silencing pathways in plants. *Trends Genet.* **22**: 268–280.
- Chen, X. 2005. microRNA biogenesis and function in plants. *FEBS Lett.* **579**: 5923–5931.
- Gascioli, V., Mallory, A.C., Bartel, D.P., and Vaucheret, H. 2005. Partially redundant functions of *Arabidopsis* DICER-like enzymes and a role for DCL4 in producing *trans*-acting siRNAs. *Curr. Biol.* **15**: 1494–1500.
- 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.
- Grundhoff, A., Sullivan, C.S., and Ganem, D. 2006. A combined computational and microarray-based approach identifies novel microRNAs encoded by human γ -herpesviruses. *RNA* **12**: 733–750.
- Gustafson, A.M., Allen, E., Givan, S., Smith, D., Carrington, J.C., and Kasschau, K.D. 2005. ASRP: The *Arabidopsis* Small RNA Project Database. *Nucleic Acids Res.* **33**: D637–D640.
- Henderson, I.R., Zhang, X., Lu, C., Johnson, L., Meyers, B.C., Green, P.J., and Jacobsen, S.E. 2006. Dissecting *Arabidopsis* DICER function in small RNA processing, gene silencing, and DNA methylation patterning. *Nat. Genet.* **38**: 721–725.
- Jones-Rhoades, M.W. and Bartel, D.P. 2004. Computational identification of plant microRNAs and their targets, including a stress-induced miRNA. *Mol. Cell* **14**: 787–799.
- Jones-Rhoades, M.W., Bartel, D.P., and Bartel, B. 2006. MicroRNAs and their regulatory roles in plants. *Annu. Rev. Plant Biol.* **57**: 19–53.
- 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 function. *Dev. Cell* **4**: 205–217.
- Kurihara, Y., Takashi, Y., and Watanabe, Y. 2006. The interaction between DCL1 and HYL1 is important for efficient and precise processing of pri-miRNA in plant microRNA biogenesis. *RNA* **12**: 206–212.
- Kusaba, M., Dwyer, K., Hendershot, J., Vrebalov, J., Nasrallah, J.B., and Nasrallah, M.E. 2001. Self-incompatibility in the genus *Arabidopsis*: Characterization of the S locus in the outcrossing *A. lyrata* and its autogamous relative *A. thaliana*. *Plant Cell* **13**: 627–643.
- Lippman, Z. and Martienssen, R. 2004. The role of RNA interference in heterochromatic silencing. *Nature* **431**: 364–370.
- 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.
- Lu, C., Tej, S.S., Luo, S., Haudenschild, C.D., Meyers, B.C., and Green, P.J. 2005. Elucidation of the small RNA component of the transcriptome. *Science* **309**: 1567–1569.
- Margulies, M., Egholm, M., Altman, W.E., Attiya, S., Bader, J.S., Bemben, L.A., Berka, J., Braverman, M.S., Chen, Y.J., Chen, Z., et al. 2005. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437**: 376–380.
- Martienssen, R.A. 2003. Maintenance of heterochromatin by RNA interference of tandem repeats. *Nat. Genet.* **35**: 213–214.
- May, B.P., Lippman, Z.B., Fang, Y., Spector, D.L., and Martienssen, R.A. 2005. Differential regulation of strand-specific transcripts from *Arabidopsis* centromeric satellite repeats. *PLoS Genet.* **1**: e79.
- Meyers, B.C., Kozik, A., Griego, A., Kuang, H., and Michelmore, R.W. 2003. Genome-wide analysis of NBS-LRR-encoding genes in *Arabidopsis*. *Plant Cell* **15**: 809–834.
- Meyers, B.C., Tej, S.S., Vu, T.H., Haudenschild, C.D., Agrawal, V., Edberg, S.B., Ghazal, H., and Decola, S. 2004. The use of MPSS for whole-genome transcriptional analysis in *Arabidopsis*. *Genome Res.* **14**: 1641–1653.
- Meyers, B.C., Souret, F.F., Lu, C., and Green, P.J. 2006. Sweating the small stuff: MicroRNA discovery in plants. *Curr. Opin. Biotechnol.* **17**: 139–146.
- Nasrallah, M.E., Liu, P., and Nasrallah, J.B. 2002. Generation of self-incompatible *Arabidopsis thaliana* by transfer of two S locus genes from *A. lyrata*. *Science* **297**: 247–249.
- Naumann, K., Fischer, A., Hofmann, I., Krauss, V., Phalke, S., Irmeler, K., Hause, G., Aurich, A.C., Dorn, R., Jenuwein, T., et al. 2005. Pivotal

- role of AtSUVH2 in heterochromatic histone methylation and gene silencing in *Arabidopsis*. *EMBO J.* **24**: 1418–1429.
- 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.
- Peragine, A., Yoshikawa, M., Wu, G., Albrecht, H.L., and Poethig, R.S. 2004. SGS3 and SGS2/SDE1/RDR6 are required for juvenile development and the production of *trans*-acting siRNAs in *Arabidopsis*. *Genes & Dev.* **18**: 2368–2379.
- Redei, G.P. 1975. *Arabidopsis* as a genetic tool. *Annu. Rev. Genet.* **9**: 111–127.
- 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.
- Rice, P., Longden, I., and Bleasby, A. 2000. EMBOSS: The European Molecular Biology Open Software Suite. *Trends Genet.* **16**: 276–277.
- Sunkar, R. and Zhu, J.K. 2004. Novel and stress-regulated microRNAs and other small RNAs from *Arabidopsis*. *Plant Cell* **16**: 2001–2019.
- Valoczi, A., Hornyik, C., Varga, N., Burgyan, J., Kauppinen, S., and 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. 2006. Post-transcriptional small RNA pathways in plants: Mechanisms and regulations. *Genes & Dev.* **20**: 759–771.
- 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.
- Vaucheret, H., Mallory, A.C., and Bartel, D.P. 2006. AGO1 homeostasis entails coexpression of MIR168 and AGO1 and preferential stabilization of miR168 by AGO1. *Mol. Cell* **22**: 129–136.
- Vazquez, F., Vaucheret, H., Rajagopalan, R., Lepers, C., Gascioli, V., Mallory, A.C., Hilbert, J.L., Bartel, D.P., and Crete, P. 2004. Endogenous *trans*-acting siRNAs regulate the accumulation of *Arabidopsis* mRNAs. *Mol. Cell* **16**: 69–79.
- Wang, X.J., Reyes, J.L., Chua, N.H., and Gaasterland, T. 2004. Prediction and identification of *Arabidopsis thaliana* microRNAs and their mRNA targets. *Genome Biol.* **5**: R65.
- Wassenegeger, M. and Krczal, G. 2006. Nomenclature and functions of RNA-directed RNA polymerases. *Trends Plant Sci.* **11**: 142–151.
- Wortman, J.R., Haas, B.J., Hannick, L.L., Smith Jr., R.K., Maiti, R., Ronning, C.M., Chan, A.P., Yu, C., Ayele, M., Whitelaw, C.A., et al. 2003. Annotation of the *Arabidopsis* genome. *Plant Physiol.* **132**: 461–468.
- 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.
- Xie, Z., Allen, E., Fahlgren, N., Calamar, A., Givan, S.A., and Carrington, J.C. 2005a. Expression of *Arabidopsis* MIRNA genes. *Plant Physiol.* **138**: 2145–2154.
- Xie, Z., Allen, E., Wilken, A., and Carrington, J.C. 2005b. DICER-LIKE 4 functions in *trans*-acting small interfering RNA biogenesis and vegetative phase change in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci.* **102**: 12984–12989.
- Yoshikawa, M., Peragine, A., Park, M.Y., and Poethig, R.S. 2005. A pathway for the biogenesis of *trans*-acting siRNAs in *Arabidopsis*. *Genes & Dev.* **19**: 2164–2175.
- Yu, D., Fan, B., MacFarlane, S.A., and Chen, Z. 2003. Analysis of the involvement of an inducible *Arabidopsis* RNA-dependent RNA polymerase in antiviral defense. *Mol. Plant Microbe Interact.* **16**: 206–216.
- Zhang, B., Pan, X., Cannon, C.H., Cobb, G.P., and Anderson, T.A. 2006. Conservation and divergence of plant microRNA genes. *Plant J.* **46**: 243–259.

Received May 19, 2006; accepted in revised form July 19, 2006.